

Search Notes

L Number	Hits	Search Text	DB	Time stamp
1	31	hiwi or piwi	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/12/13 18:18
2	19	(hiwi or piwi) and human	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/12/13 18:18
3	8	(hiwi or piwi) same human	USPAT; US-PGPUB; EPO; JPO; DERWENT	2003/12/13 18:18

Set	Items	Description
S1	220	HIWI OR PIWI
S2	95	RD (unique items)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 19:08:13 ; Search time 47 Seconds
(without alignments)
2617.300 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 4102

Sequence: 1 MIFGVNTRQLDHWKSKTG.....VQSIHREPNLSNRLYYL 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq 19Jun03.*

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4061	99.0	861	21	AA190235 Human piwi protein
2	3957	96.5	862	21	AA190234 Mouse piwi protein
3	2764	67.4	523	23	ABP67995 Human colon cancer
4	1468	35.8	866	22	ABB62102 Drosophila melanog
5	1428.5	34.8	843	22	ABB62084 Drosophila melanog
6	1416.5	34.5	843	21	AA190233 Drosophila piwi pr
7	1389.5	33.9	580	22	AAU07866 Polypeptide sequen
8	1324.5	32.3	498	23	ABP43865 Human mRNA sequen
9	1230	30.0	530	22	AA194209 Human protein sequ

10	962	23.5	371	22	AA192722 Human protein sequ
11	652	15.9	130	21	AB158977 Breast and ovarian
12	580	14.1	341	22	AB158957 Novel human diagn
13	510	12.4	868	21	AA192119 Arabidopsis thalia
14	510	12.4	1048	21	AA192118 Arabidopsis thalia
15	508.5	12.4	859	23	ABP64718 Human protein SQ
16	508.5	12.4	860	23	ABG97473 Human NAAP7 from I
17	503	12.3	884	21	AA192594 Arabidopsis thalia
18	503	12.3	906	21	AA192593 Arabidopsis thalia
19	503	12.3	983	21	AA192592 Arabidopsis thalia
20	503	12.3	988	21	AA192591 Arabidopsis thalia
21	503	12.3	1064	21	AA192590 Arabidopsis thalia
22	500	12.2	950	22	ABG67046 Drosophila melanog
23	500	12.2	984	22	ABG62573 Drosophila melanog
24	500	12.2	984	22	ABG67045 Drosophila melanog
25	500	12.2	984	23	ABG70016 Larval viability a
26	497	12.1	860	22	AA190292 Human protein SQ
27	497	12.1	860	22	AA194291 Human protein sequ
28	497	12.1	860	24	ABG72612 Human cytokine or
29	496	12.1	860	24	ABU11762 Human MDR1 polyrep
30	485.5	11.8	820	21	AA192120 Arabidopsis thalia
31	485.5	11.8	836	21	AA192595 Arabidopsis thalia
32	473.5	11.5	782	22	AA193139 Human protein sequ
33	427	10.4	962	22	ABG63278 Drosophila melanog
34	417.5	10.2	1145	22	AA190770 Drosophila melanog
35	372	9.1	938	22	AA192553 Neurospora crassa
36	340.5	8.3	199	22	ABG05956 Novel human diagn
37	321.5	7.8	326	21	AA192610 Arabidopsis thalia
38	312	7.6	1020	22	AAU01856 C. elegans RNA int
39	303	7.4	108	22	AA004945 Human polypeptide
40	301	7.3	140	22	ABG12914 Novel human diagn
41	297.5	7.3	406	21	AA192599 Zea mays protein f
42	292	7.1	227	21	AA192611 Arabidopsis thalia
43	290	7.1	252	22	AA197112 Translation initia
44	251.5	6.1	352	21	AA192551 Zea mays protein f
45	251.5	6.1	353	21	AA192550 Zea mays protein f

ALIGNMENTS

RESULT 1

AA190235

ID AA190235 standard; Protein; 861 AA.

XX AC AA190235;

XX DT 29-AUG-2000 (first entry)

XX DE Human piwi protein, designated hiwi.

XX KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;

XX KW anaemia; immunodeficiency; male infertility; human.

XX OS Homo sapiens.

XX EH Key Location/Qualifiers

XX FT Misc-difference 76

XX FT /label= Leu, Ile

XX FT /note= "encoded by NTA"

XX FT Misc-difference 303

XX FT /label= Leu, Ile

XX FT /note= "encoded by NTA"

XX FT Misc-difference 735

XX FT /label= Leu, Ile

XX FT /note= "encoded by NTA"

XX FT W0200032039-A1.

XX PD 08-JUN-2000.

XX PF 03-DEC-1999;

XX 99WO-US28764.

```

PR 04-DEC-1998; 98US-0110901.
XX (UYDU-) UNIV DUKE.
XX Lin H;
XX WPI; 2000-412085/35.
XX N-PSDB; AAA07588.
XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
XX therapy of diseases such as cancer and in various research and
XX diagnostic applications -
XX Claim 4; Page 189-194; 201pp; English.
XX This sequence represents the human piwi family protein, designated
XX hiwi. The piwi family nucleic acids and polypeptides are used in gene
XX therapy of diseases such as cancer and also in various research and
XX diagnostic applications. The sequences can also be used to treat
XX tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX OS Sequence 861 AA;
SQ Query Match 99.0%; Score 4061; DB 21; Length 861;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 768; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GVNTRQNLHVKSCTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARLRSL 63
Db 90 GVNTRQNLHVKSCTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARLRSL 149
QY 64 LFOHEDLIGKCHAFDGTILFLPKRLQOKVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 123
Db 150 LFOHEDLIGKCHAFDGTILFLPKRLQOKVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 209
QY 124 QFYNIIFRLLKIMNLOQIGRNYNPNDPIDIPSHRLVWPGETTSIIQYENSIMLCTDV 183
Db 210 QFYNIIFRLLKIMNLOQIGRNYNPNDPIDIPSHRLVWPGETTSIIQYENSIMLCTDV 269
QY 184 SHKVLRSVTLDFMFNFYHQTEEHKFOQVSKELIGLVLTIKYNNKTYRVDIDWDQNP 243
Db 270 SHKVLRSVTLDFMFNFYHQTEEHKFOQVSKELIGLVLTIKYNNKTYRVDIDWDQNP 329
QY 244 STFKKADSEVSFLYRKQYNOEITDLKQPLVVSQPKRRRPGGTLPGLPAMLIPELCYL 303
Db 330 STFKKADSGVSFLYRKQYNOEITDLKQPLVVSQPKRRRPGGTLPGLPAMLIPELCYL 389
QY 304 TGLTDKMNDFNMKDLAVHTLTPQQRQREYGRLLIDYTHKNDVORELRDGLSFDNL 363
Db 390 TGLTDKMNDFNMKDLAVHTLTPQQRQREYGRLLIDYTHKNDVORELRDGLSFDNL 449
QY 364 LSPSGRILOTEKHGGKTFDNPQFADWSKETRGAPLISVKPLDNWLLIYTRRYEAA 423
Db 450 LSPSGRILOTEKHGGKTFDNPQFADWSKETRGAPLISVKPLDNWLLIYTRRYEAA 509
QY 424 SLIQLNFKVTPANGQMKAMIEVDDETEAVLRVLOKVADTQIVVCLSSNKKDYD 483
Db 510 SLIQLNFKVTPANGQMKAMIEVDDETEAVLRVLOKVADTQIVVCLSSNKKDYD 569
QY 484 AIKKYLCTDCTPSCQVAVTLGKQTVMAIATKIALQNMCKMGELWRVDIPLKLVMI 543
Db 570 AIKKYLCTDCTPSCQVAVTLGKQTVMAIATKIALQNMCKMGELWRVDIPLKLVMI 629
QY 544 GIDCHDMTAGRRSITAGFVASINEGTRWFSRIFQDRGOELVDGLKVLQALRAMNSC 603
Db 630 GIDCHDMTAGRRSITAGFVASINEGTRWFSRIFQDRGOELVDGLKVLQALRAMNSC 689
QY 604 NEYMERIIIVYRDGVDGOLKTLVNYEVPQFLDCLKSIQRGNPRLTVIVVKRVTNTRFF 663
Db 690 NEYMERIIIVYRDGVDGOLKTLVNYEVPQFLDCLKSIQRGNPRLTVIVVKRVTNTRFF 749
QY 664 AQSGRLQNLPGTVIDVEVTRPEWYDFIIVSQAVRSGSVSPTHYVNVYDNSGLKPDHIQ 723

```

```

Db 750 AQSGRLQNLPGTVIDVEVTRPEWYDFIIVSQAVRSGSVSPTHYVNVYDNSGLKPDHIQ 809
QY 724 RLTYKLCIIYNNWPGVIRVPAPCOYAHKLAFLVQGSIHREENLSLSNRLYYL 775
Db 810 RLTYKLCIIYNNWPGVIRVPAPCOYAHKLAFLVQGSIHREENLSLSNRLYYL 861

RESULT 2
AAAY90234
ID AAAY90234 standard; Protein; 862 AA.
XX AC AAAY90234;
XX DT 29-AUG-2000 (first entry)
XX DE Mouse piwi protein, designated miwi.
XX KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
XX anaemia; immunodeficiency; male infertility; mouse; ds.
XX OS Mus sp.
XX FH Key Location/Qualifiers
FT Misc-difference 90 /label= Leu, Ile
FT /note= "encoded by NNT"
FT Misc-difference 216 /note= "unspecified amino acid; encoded by NTC"
FT Misc-difference 383 /label= Leu, Ile
FT /note= "encoded by NTC"
FT Misc-difference 816 /label= Leu, Ile
FT /note= "encoded by NTC"
XX WO2000032039-A1.
XX PD 08-JUN-2000.
XX PF 03-DEC-1999; 99WO-US28764.
XX PR 04-DEC-1998; 98US-0110901.
XX PA (UYDU-) UNIV DUKE.
XX Lin H;
XX WPI; 2000-412085/35.
XX N-PSDB; AAA07587.
XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
XX therapy of diseases such as cancer and in various research and
XX diagnostic applications -
XX Claim 4; Page 180-185; 201pp; English.
XX This sequence represents the mouse piwi family protein, designated
XX miwi. The piwi family nucleic acids and polypeptides are used in gene
XX therapy of diseases such as cancer and also in various research and
XX diagnostic applications. The sequences can also be used to treat
XX tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX OS Sequence 862 AA;
SQ Query Match 96.5%; Score 3957; DB 21; Length 862;
Best Local Similarity 95.9%; Pred. No. 0;
Matches 740; Conservative 19; Mismatches 13; Indels 0; Gaps 0;

QY 4 GVNTRQNLHVKSCTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARLRSL 63
Db 91 GVNTRQNLHVKSCTGSSGIIVRLSTNHFRLTSRPQWALYQYHIDYNPLMEARLRSL 150
QY 64 LFOHEDLIGKCHAFDGTILFLPKRLQOKVTEVFSKTRNGEDVRITITLTNELPPTSPTCL 123

```

Db 151 LFOHEDLIARHAFDGTILFLPKLQHKVTEVFSQTEGHEVRITITLTNELPPTSPCL 210
Qy 124 QFYNIILFRLKIMNLOIGRNYNPNPDIDIPSHRLVIWPGFTTSILQYENSIMLCTDV 183
Db 211 QFYNIIFRLLKIMNLOIGRNYNPNPDIDIPSHRLVIWPGFTTSILQYENNIMLCTDV 270
Qy 184 SHKULRSVTLDFMFNFYHOTEHFKFOEQVSKELIGLVLTYYNNKTYRVDIDWDQPK 243
Db 271 SHKULRSVTLDFMFNLVQOTEHFKFOEQVSKELIGLVLTYYNNKTYRVDIDWDQPK 330
Qy 244 STFKKADGSEVSFLYYRKQYNQIEITDLKQPVLSQPKRRRPGGTLPFAPMLIPELCYL 303
Db 331 STFKKADGSEVSFLYYRKQYNQIEITDLKQPVLSQPKRRRPGGTLPFAPMLIPELCYL 390
Qy 304 TGLTDKMRDNFNMKDLAVHTLTPEQOREVGRLLDIYHKNDNVQRELDRWGLSFDNSL 363
Db 391 TGLTDKMRDNFNMKDLAVHTLTPEQOREVGRLLDIYHKNDNVQRELDRWGLSFDNSL 450
Qy 364 LSFSGRIILQTEKHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRRYEAA 423
Db 451 LSFSGRIILQSEKHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRRYEAA 510
Qy 424 SLIQNLFKVTPAMQMKRKAIMEVDRTPEAYLRLVLOQKVTDATQIVVCLLSSNRKDYD 483
Db 511 SLIQNLFKVTPAMQMKRKAIMEVDRTPEAYLRLVLOQKVTDATQIVVCLLSSNRKDYD 570
Qy 484 ALKKYLCTDPTSPQCVVARTLGKQOTVMAIATKIALQNMCKMGELWRVDIPKLVMIV 543
Db 571 ALKKYLCTDPTSPQCVVARTLGKQOTVMAIATKIALQNMCKMGELWRVDIPKLVMIV 630
Qy 544 GIDCVHDMTAGRSIAGFVASINEGTRWFSCIFQDRGOELVDGLKVCLOALRAWNSC 603
Db 631 GIDCVHDMTAGRSIAGFVASINEGTRWFSCIFQDRGOELVDGLKVCLOALRAWNSC 690
Qy 604 NEYMPRIIVYRDGVGDGQKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKRYNTRFF 663
Db 691 NEYMPRIIVYRDGVGDGQKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKRYNTRFF 750
Qy 664 AQSGELQNLPGTVDIVETPEWDFPIVQAVRSGSVSTHYNVIYDNSGLKPDHIQ 723
Db 751 AQSGELQNLPGTVDIVETPEWDFPIVQAVRSGSVSTHYNVIYDNSGLKPDHIQ 810
Qy 724 RLTYKLCCHYYNWPVIRVAPCOYAHKLAFLVGSIHREPNLSNRLYYL 775
Db 811 RLTYKLCCHYYNWPVIRVAPCOYAHKLAFLVGSIHREPNLSNRLYYL 862

RESULT 3
ID ABP67995 standard; Protein; 523 AA.
XX
AC ABP67995;
XX
DT 13-DEC-2002 (first entry)
XX
DE Human colon cancer related polypeptide SEQ ID NO 2603.
XX
KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine.
XX
OS Homo sapiens.
XX
PN WO200258534-A2.
XX
PD 01-AUG-2002.
XX
PF 19-NOV-2001; 2001WO-43704.
XX
PR 20-NOV-2000; 2000US-252222P.
PR 06-FEB-2001; 2001US-267011P.
PR 28-MAR-2001; 2001US-279670P.
PR 10-JUL-2001; 2001US-304037P.
XX

(CORI-) CORIXA CORP.
Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GB;
WPI; 2002-608400/65.
N-PSDB; ABV89287.
New isolated tumor colon polynucleotide and polypeptide, useful for the
diagnosis, prevention and/or treatment of cancer, in particular colon
cancer
Claim 2; SEQ ID NO 2603; 266pp + Sequence Listing; English.
The invention relates to a human colon tumour expressed polynucleotide
(I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
sequences that hybridize to (i), under moderately stringent conditions;
(v) sequences having at least 75% or 90% identity to (i); or (vi)
degenerate variants of (i). The compositions and methods of the present
invention are useful for the diagnosis, prevention and/or treatment of
cancer, particularly colon cancer. (i) can be used in gene therapy and
(II) and (II) are useful in pharmaceutical compositions such as vaccines.
Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
Query Match 67.4%; Score 2764; DB 23; Length 523;
Best Local Similarity 100.0%; Pred. No. 9.7e-261;
Matches 522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 254 VSFLEYRKYQNOEITDLKQPVLSQPKRRRPGGTLPFAPMLIPELCYLGLTDMRND 313
Db 2 VSFLEYRKYQNOEITDLKQPVLSQPKRRRPGGTLPFAPMLIPELCYLGLTDMRND 61
Qy 314 FNVKMDLAVHRLTPEQOREVGRLLDIYHKNDNVQRELDRWGLSFDNSLSFSGRILO 373
Db 62 FNVKMDLAVHRLTPEQOREVGRLLDIYHKNDNVQRELDRWGLSFDNSLSFSGRILO 121
Qy 374 EKHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRRYEAA NSLIQNLFKVT 433
Db 122 EKHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRRYEAA NSLIQNLFKVT 181
Qy 434 PAMGQMKRKAIMEVDRTPEAYLRLVLOQKVTDATQIVVCLLSSNRKDYDAIKKYLCTDC 493
Db 182 PAMGQMKRKAIMEVDRTPEAYLRLVLOQKVTDATQIVVCLLSSNRKDYDAIKKYLCTDC 241
Qy 494 PTPSQCVVARTLGKQOTVMAIATKIALQNMCKMGELWRVDIPKLVMIVGDCYHDMTA 553
Db 242 PTPSQCVVARTLGKQOTVMAIATKIALQNMCKMGELWRVDIPKLVMIVGDCYHDMTA 301
Qy 554 GRSIAGFVASINEGTRWFSCIFQDRGOELVDGLKVCLOALRAWNSCNEYMPRIIV 613
Db 302 GRSIAGFVASINEGTRWFSCIFQDRGOELVDGLKVCLOALRAWNSCNEYMPRIIV 361
Qy 614 YRDGVGDGQKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKRYNTRFFAQSGRLQNP 673
Db 362 YRDGVGDGQKTLVNYEVPQFLDCLKSIGRGYNPRLTVIVVKRYNTRFFAQSGRLQNP 421
Qy 674 LPGTVIDVEVTRPEWDFPIVQAVRSGSVSTHYNVIYDNSGLKPDHIQRLTYKLCCHY 733
Db 422 LPGTVIDVEVTRPEWDFPIVQAVRSGSVSTHYNVIYDNSGLKPDHIQRLTYKLCCHY 481
Qy 734 YNWPVIRVAPCOYAHKLAFLVGSIHREPNLSNRLYYL 775
Db 482 YNWPVIRVAPCOYAHKLAFLVGSIHREPNLSNRLYYL 523

RESULT 4
ID ABP62102 standard; Protein; 866 AA.

XX ABB62102;
 XX 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 13098.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical.
 XX Drosophila melanogaster.
 XX WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 XX N-PSDB; ABL06205.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions -
 XX Disclosure; SEQ ID NO 13098; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 XX useful in developmental biology and in elucidating cell signalling and
 XX cell-cell interactions in higher eukaryotes for the development of
 XX insecticides, therapeutics and pharmaceutical drugs. The invention
 XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 XX sequences (ABL01840-ABL16175) and the encoded proteins
 XX (ABB57737-ABB72072).
 XX The sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 866 AA;
 XX Query Match 35.8%; Score 1468; DB 22; Length 866;
 XX Best Local Similarity 39.4%; Pred. No. 1.2e-133;
 XX Matches 305; Conservative 154; Mismatches 291; Indels 24; Gaps 12;
 XX 17 SKTGSSGIIVRLSTNHFRLTSRPOWALYOHIDYNPLMEARRLSALLFQHEGLICKCHA 76
 XX 102 SKGVGVGTHITVQANFKVLRNWTIYQYRVDFTDVEATRLRRFLYEHKGLGG-VI 160
 XX 77 FPGTILFLP--KRLQOK--VTEVFKTRNGEDVRITITLNLPLPTSPCTCLQFYNIIFR 131
 XX 161 FPGTNNFCINQKFAVDSPVLELVTKSRAGENIEIKIKAVGSVQSTDAEQFQVNLILR 220
 XX 132 RLKIKINMLQOIGRNYNPNPDIDIPSHRLVWPGFTTSILOYENSIMLCTDYSHKVLRS 191
 XX 221 RAMEGDLKLVSRYYDPQAKINLENFRMLQWPGYQTSIRQHENIDILLCEICHKVMRT 280
 XX 192 TVLDFMNFVHOPEEHKFBQSVSKELIGLVILKYNKTVRVDIDWDQNKSTFKKADG 251
 XX 281 TLYNLSDAIRDSD--YQSTFRVANGWVILTDYNNKTRIDDDVQSTPLCKFTNDG 338
 XX 252 SEVSFLYKRYKQYNOBITDLKQPLVLSQPKRRPGGTLFGPAMLIPELCYLTGLTKDKMR 311
 XX 339 -EISYVDYIKRYNIIRDLKQPLVMSRPTDKNIRGN-DQAIMIPELARATGMDAMR 396
 XX 312 NPEWMDLAVHTLPTPEQRQREVRGLIDIYHKNDNVQRELWDGLSFDNSLLSFSGRIL 371

Db 397 ADFTLRAMSEHTRINPDRIERLRFMFKRLKSKQSVETLTKSMNIELDSALVEIPARVL 456
 QY 372 QTEKIHQGGKTFDYNPQFADWSKETRGAPLISVPLDNWLLIYTRRNYEAANSLQNLK 431
 Db 457 PPEKILFGNOKIFVCDARADWNEFRCTSMFKVHNIRWYVITPSRNLRETQEFVQMCIR 516
 QY 432 VTPAMGOMRKAIMIEV--DDTEAYLRVLOOKTADTQIVVCLLSSNRKDKYDAIKKYL 490
 Db 517 TASSMKMNICPIYBEIPDDRNGTYSQAIDNAAANDPQIVMVMRSPNEEKYSCKIKRT 576
 QY 491 TDCPTPSQCVVARTILGKQOT---VMAITATIALOMCKMKGELWRVDIPLKLVIMVGD 546
 Db 577 VDRFVPSQVVTLKVIAPRQKPTGLMSTATKVQMNAKMGAPQVQVIFLHGLMTVGPD 636
 QY 547 CVHDMTAGRRSIAGFVASINEGMT--RWFSCIFQDRGOELVDGLKVCILQALRAWNSNE 605
 Db 637 VCHSPKKNKSYGAFVATMDQKESFRYFSTVNEHIKGOELSEQMSVNVACALRSYQEOHR 696
 QY 606 YNPSRIIVRDGVGDGQKTLVNYEVPQFLDCL----KSIGRGYNPRLTVIVVKRVNTR 661
 Db 697 SLPERILFRDGVGDGQLYQVNVSEVNTLKDRLDIYKSAGKQEGCRMTFIIVSKRINSR 756
 QY 662 PFAQSGGRLQNLPGCTVIDVEVTRPEWYDFRIVSOAVRSGVSUPHYNVYDNSGLKPDH 721
 Db 757 YF--TGHR--NPVPGTVVDDVITLPERYDFFLVSOAVRIGTVSPTSYNVISDNMGLNADK 812
 QY 722 IQRITKYLCHYIYNNPVGIVRVPAPQYAHKLAFLVQSIHREPNSLSNRLYYL 775
 Db 813 LQMLSYKMYNYNSGIVRVPVAVCHYAKLAFLVAESINRAPSAGLQNLQYFL 866
 RESULT 5
 ABB62084
 ID ABB62084 standard; Protein; 843 AA.
 XX ABB62084;
 XX 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 13044.
 XX Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical.
 XX Drosophila melanogaster.
 XX WO200171042-A2.
 XX 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 XX N-PSDB; ABL06187.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 XX genes from Drosophila and for elucidating cell signalling and cell-cell
 XX interactions -
 XX Disclosure; SEQ ID NO 13044; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 XX useful in developmental biology and in elucidating cell signalling and
 XX cell-cell interactions in higher eukaryotes for the development of


```

317 MKDLAVTRLTLPBORQREVGHLIDYIHKNDVQ---RELRDWGLSFDNLLSFSGRILQT 370
384 MRAMSSYTRMNPQR---TDLRAFNFHRLQNTPPSVKVLRDWNMELDKNVEVQGRILQ 440
374 EKI--HGGKTFDYNPQADKSKETRGAFLISVKP--LDNMLLIYTRNEYEAANSLIQNL 429
441 QNIVFHNGKVPAGEN---ADQRHFRDQRMJLTPSDGLDRWAVIAPQRNSHELRTLLDLS 497
430 FKVTPAMGMQMRK-AIMIEVDRTDEAYLRVLQCKVTADTQIVVCLLSSNRKDKYDAIKKY 488
498 YRAASGMGLAIRSPQEFIIYDRGTGYVRAMDDCVSRDPKLLILCLVPNDNAERYSSIKR 557
489 LCTDCTPTSCQVARTIGCQOTVMAIAIKIALQNMCKMGGELWRVDIPLKLVMIIGIDCY 548
558 GYVDRAVPTQVVTLLTKT-KNRSLSMSIATKIALQNLCKLGYPFWMIELPLSLGLMTTGFDA 616
549 HDMTAGRRSIAGFVASIN-EGMTWFSRCIFQDRGQELVDGILKVCLOAALRAWNSCNEYM 607
617 KSTRDRKAYGALLASMDLQNSYFSTVTECSAFDVLANTLWPMIAKALROYQHEHRKL 676
608 PSRIIVYRDGVDGQLKTLVNYEVPQFDCLKLSIGRGY-----NPRLTIVIVKKRVNTR 661
677 PSRIIVFYRDGVSSGSLKQLFEFEVKDIIETKLT---EYARVQLSPQLAYIVVTRSMNTR 733
662 FFAQSGGELQNLPGTVIDVEVTRPEWYDFPIVSCAVRSQVSPHYNVIVDNSGLKPDH 721
734 FFLNG----QNPPPTGIVDDITLPERYDFYLVSQQVRQGTVPSTSYNVLYSSMGXSPEK 789
722 IQRITYKLCHYYNWPVIRVPAPCOYAHKLAFLVGQSIHREPNISSLNRLVYL 775
790 MQKLTYKMHCHLYYNSGTRVPVAVCOYAKKATLVGTNLHLSIPQNALEKKFYIL 843

RESULT 7
AAU07866
ID AAU07866 standard; Protein; 580 AA.
XX
AC AAU07866;
XX
DT 18-DEC-2001 (first entry)
XX
DE Polypeptide sequence for mammalian Spg16.
XX
KW Mammalian; reproductive-specific protein; male infertility;
KW spermatogenesis; sperm count disorder; anti infertility; reproduction.
XX
OS Mammalia.
XX
FN WO200166752-A2.
XX
PD 13-SEP-2001.
XX
PF 07-MAR-2001; 2001WO-US07371.
XX
XX
PR 07-MAR-2000; 2000US-0187518.
XX
PR 12-JAN-2001; 2001US-0261557.
XX
PA (WHEED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Wang PJ, Page DC;
XX
XX WPI; 2001-570774/64.
XX
DR N-PSDE; AAS13630.
DR

```

reproductive-specific proteins are useful for diagnosing infertility which is a result of reduced sperm count, reduced sperm motility, malformed sperm or combinations of these. The sequences of the invention are useful as markers for spermatogonial cells, for identifying genes or proteins characteristic of male infertility, diagnosing or aiding in the diagnosis of infertility in men, and for contraception in which sperm production or sperm count is reduced or defective sperm is produced. Antibodies to reproductive-specific proteins are useful for determining the presence of these proteins in a sample obtained from a man being assessed for infertility, for identifying the expression of genes in particular cell type or particular developmental stage, for studies of spermatogenesis, and for immunofluorescence of germ cells or in Western blots for assessing the presence of the protein the antibody binds. The sequences of the invention are also useful for treating disorders of reduced sperm count, and for increasing sperm count and/or sperm activity. The nucleic acids of the invention are useful in gene therapy. AAU07859-AAU07882 represent the mammalian reproduction-specific proteins of the present invention.

AA	Sequence	580 AA;
SQ		

Query Match	33.9%;	Score 1389.5;	DB 22;	Length 580;
Best Local Similarity	44.2%;	Fred. No. 3e-126;		
Matches 258;	Conservative 118;	Mismatches 199;	Indels 9;	Gaps 8;

Qy		197	MNFYFQTGTEHKBQBOVSKELIGLVLTKYNKTYRVDDDWDQNPKSTFKADSEVSF	256
Db		1	MHAIIYQONKEH-FQDECSKLVGSIIVITRYNNRTYRDDVDWNKPDKDSFYMSDGKETIF	59
Qy		257	LEYRYOYNQOEITDLQPVLVSOP-KERRPGGGTLPGPAMLIPELCYLTLGLTDKMNDEN	315
Db		60	LEYYSKNYGITVEKDDOPLIHIHPSEKQNNHGMLLKGEILLPPLSPFWTGIPKXWKDOFR	119
Qy		316	VMKDLAVHTRLTPEQRQREVRGLIDYTHKNDNVQRELDMGLSDSNLLSFSGRILQTEK	375
Db		120	AKMDLTQQINLSFKOHGALECLLQRISQNETASNELTRWGLSLHKDVHKIEGRLPMER	179
Qy		376	IHOGGKTEDYNPOPADWSKE-TRGAPLISVKPLDNWLIIYYTRNVEAANSILIONLFKVTP	434
Db		180	INLRNTSF-VTSEGLNAWKEVTEDASILTI-PMHFALFYFKRAMDQARELVNMLEKIAG	237
Qy		435	AMGQMOKEAIMIEV-DDRTEAYLRVLQ--KYVADTOIVVCLLSSNRKDKYDAIKVLYCT	491
Db		238	PICMRYSPPAWVELKDRIETIYIRTIQSILGVBEKIOMVVCIIMGTRDDLGAIKKLCCV	297
Qy		492	DCTPTSCVVARILKGOOTWMATKIALQANCKMGSELWRVIDPLKLVMIVGIDCYDHM	551
Db		298	QSVPSQVINVRTIGOPTLRSVAQKILLQMNCKLGGELNGVDPLKQLMWIGNDVYHD	357
Qy		552	TAGRSTIAGVASINEGWTFWGRSCPIQDRGOELVDGLKVCLOQAALRAWNSCNEVMPERI	611
Db		358	SRCWRSVVGFGVASINTLTKWYSRVFPQPHQEIVDSLKLGLVGS�KKYEEVNHHCLPEKI	417
Qy		612	IYVRDGVGDQLKTLVNYEVPQFLDCLKSIGRGYNRPRLTVIWWKRVNTRFPAOSGGRLQ	671
Db		418	VYTRDGVSGQLKTVANIETIQOKCEAFD-NYHPKXWVFVVOQKISTNIYLLAADHFV	476
Qy		672	NPLFGTVIDYEVRPEWYDFFIYSQAVRSQSVSPHYNNVIYDNSGLKPDHTRLYTKLCH	731
Db		477	TPSPGTIVDHTITSCEWVD FYLLAAHVROCGGPIPTHYICVLNTANISPDMHQRLTFKLCH	536
Qy		732	IYNNWPGVTRVPACQVAHKLAFIVGOSITHREPNLISNRUYL	775
Db		537	MYNNWGTTTRVPACKYAHKLAFSLGQILHLHEPAIQLCGNLFFL	580

RESULT 8

ABP43865
ID ABP43865 standard: Protein: 498 AA:

AC ABP43865:

DT 26-FEB-2003 (first entry)

XX Human mRNA sequence.

XX Neuroprotective; immunomodulator; cancer;

KW cytosolic; anti-inflammatory; gene therapy; nutritional supplement;

KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;

KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;

KW vulnary.

XX Homo sapiens.

XX WO200231111-A2.

XX 18-APR-2002.

XX 11-OCT-2001; 2001WO-US27760.

XX 12-OCT-2000; 2000US-0687527.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI; 2002-426278/45.

XX N-PSDB; AB061109.

XX New polypeptides and their encoded proteins, useful as nutritional

PT sources or supplements, or in gene therapy, particularly for treating

PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or

PT inflammation -

XX Claim 20; SEQ ID # 768; 357pp + sequence listing; English.

XX The invention relates to 446 newly isolated polynucleotide sequences.

CC The activity of polynucleotides of the invention may be described as,

CC vulnary, neuroprotective, immunomodulator, cytosolic and

CC anti-inflammatory. Compositions comprising nucleic acids of the invention

CC are useful for treating a mammalian subject, or as nutritional sources or

CC supplements. These are useful in gene therapy, particularly for treating

CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,

CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or

CC inflammation. The nucleic acids and polypeptides are also useful in

CC diagnostic and research methods. The sequences, given in records

CC ABP4354-ABP4398 represent polypeptides encoded by polynucleotides of

CC the invention.

CC NOTE: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 498 AA;

SQ

Query Match 32.3%; Score 1324.5; DB 23; Length 498;

Best Local Similarity 51.4%; Pred. No. 5.3e-120;

Matches 259; Conservative 86; Mismatches 150; Indels 9; Gaps 4;

QY 275 VLVQPKRGEGTLPAGMILPELCYLTGTDKMRDNFNMKDLAVHTRLTPEQRQE 334

DB 1 MLVSLKKKRNDSN-PQLAHLPELCFTLTQATSDPQLMKAVAEKTRLSFGQRQR 59

QY 335 VGLIDYIHNDVQRELDRWGLSFDNLSLFSGRILQTEKHQGGKTFDYNPQ---FAD 391

DB 60 LARLVNIOQNTWARTETWGLHFGSQ-ISLGRIVPSEKILMQ----DHICQPSAAD 114

QY 392 WSKETRGAFLISVKPLDNMLLIYTRRNYEAANSLIQNLFKVTMPQMGRKAIMIEVD 451

DB 115 WSKDIRTKILNAQSLNTWLILCSDRTEYVAESFLNCLRRVAGSMGFNDYPKIIVQEN 174

QY 452 TEAYLRVLOQKVTADTQIVVCLLSNRKOKYDAIKYLCIDCTDCTPQCQVARTLHGQQT 511

DB 175 PRAFVRAIQVYDPDVLQVLMCILFNSQKTYDSIKYLSDDCPVSPQCQVARTLNLQGM 234

QY 512 MATATKIALQNMCKMGELWRVDIPLKLVMIIVGIDCYHDMTAGRRSTAGFVASINSGMTR 571

DB 235 MSATKIAMQMTCKLGGELWAVEIPKLSMIVGVGIDVCKDALSCKDVMVGCVASVNPTRITR 294

QY 572 WFSRCIFQDRQOELVDGLKVCLOAALRAVNSCNEYMPSRIIVYRDVCGDGLKTLVNYEV 631

DB 295 WFSRCILQRTWTDVADCLKVKFMTGALNKWYKYNHDLPARIIIVYRAGVGDGOLKTIIEV 354

QY 632 PQFLDCLKSIGRGYNPRLTVIWKRVNTRFFAOSGGRLOQNPPLGCTVIDVETRPEWYDF 691

DB 355 PQLSSVAESSNTSSRLSVIVRKKCMRPFTEHNRVQNPPLGTVVDSEATRNWYDF 414

QY 692 FIVSQAVRSGSVSPHYNNYVYDNSGLKPDHQRITLYKLCIHYNNWPGVIRVPAPCOYAHK 751

DB 415 YLISQVACRGTVSPYYNVIYDNGLKFEDHMQRLTFKLCHLYYNNWPGVIRVPAPCOYAHK 474

QY 752 IAFVQGSIHREPNSLSNRLYYL 775

DB 475 LTFIGAQSIIHKEPSLELANHLFYL 498

RESULT 9

AB94209

ID AAB94209 standard; Protein; 530 AA.

XX AAB94209;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:14559.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX Claim 8; SEQ ID 14559; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

OY 767 SLSNRLXYL 775
 DB 363 QLCENLFFL 371
 RESULT 11
 ID AAB58977
 XX AAB58977 standard; Protein; 130 AA.
 AC AAB58977;
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Breast and ovarian cancer associated antigen protein sequence SEQ ID 685.
 XX
 KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antineoplastic; antitumor; antitumor; antitumor; antitumor;
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200055173-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05881.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM;
 XX
 DR WPI; 2000-611515/58.
 DR N-PSDB; AAF21880.
 XX
 PT New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 PS
 PS Claim 11; Page 1139; 1299pp; English.
 CC
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antineoplastic; antitumor; antitumor; antitumor;
 CC antibacterial; antifungal; antiparasitic and cardiant activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemia; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 CC
 SQ Sequence 130 AA;
 Query Match 15.9%; Score 652; DB 21; Length 130;
 Best Local Similarity 99.2%; Pred. No. 3.2e-55;
 Matches 123; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 492 DCPTSCQCVARTLKGKOOTVMAIAIKIALQWCKMGGLRWVDIPLKLVMIYIGIDCYHDM 551
 DB 5 DCPTSCQCVARTLKGKOOTVMAIAIKIALQWCKMGGLRWVDIPLKLVMIYIGIDCYHDM 64
 OY 552 TAGRSIAGFVASINEGTRWFSRCIFODRGOELVDGLKVCIAALRAWNSCNEYMPRSI 611
 DB 65 TAGRSIAGFVASINEGTRWFSRCIFODRGOELVDGLKVCIAALRAWNSCNEYMPRSI 124
 OY 612 IVYR 615
 DB 125 IVYR 128
 RESULT 12
 AAG05957
 ID AAG05957 standard; Protein; 341 AA.
 XX
 AC AAG05957;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #5948.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS70144.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PT
 PT Claim 20; SEQ ID No 36316; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAG0010-AAG0377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

764 PNLSLSNRL 772

Db 818 PETSDBGSM 826

RESULT 14
AAG42118
ID AAG42118 standard; Protein; 1048 AA.
XX
AC AAG42118;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 52487.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN BP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139942.
PR 18-JUN-1999; 99US-0139943.
PR 18-JUN-1999; 99US-0139944.
PR 18-JUN-1999; 99US-0139945.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	23-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158222.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160747.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161922.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.
Query Match 12.4%; Score 510; DB 21; Length 1048;		
Best Local Similarity 25.1%; Pred. No. 8,2e-40;		
Matches 213; Conservative 136; Mismatches 376; Indels 124; Gaps 33;		
Qy	20 GSSGHIVRLSTNHFRILTRSPQWALQYHYHIDYNPLMEAPRLRSALLFOHEDLIGKH----	75
Db	186 GOSGRCICVKAHNF-FAELPKDLHHYDVITTEPVSIGVNAVMKQLVDNVRDSHLGSR 244	
Qy	76 --AFPG-----TILFIP-----KLOQKTVIEVFSKTRNGEDVRITI 109	
Db	245 LPAYDGRKSLYTAGLPFPNSKEFRINLLDEEVGAGGQRRREFKVIKLVARADLHLGLM 304	

QY	110	TLTNELPPTSPICLOFYNIIFRELLKIMNLQOLGRNVYNENDPIDIPSHR-----LVWLP	166
DB	305	FLEGKSDAPQBALQVLDIVLRE-LFTSYVIPGRSFYSP-----DIKKKSLGDLGSWR	359
QY	165	GFTTSLQIYENSIMLCTDVSHKV-LRSETVLDFWFPYHQTEEHK-----FOEQVSKELI	218
DB	360	GFYQSIRPTQMGLSLIDMSSTAFIEANPVIQFCDLLNREISRPPLSDADRVKIKKALR	419
QY	219	GL-VVLTKYNN--KTYRVDDIDWDQNPKFTFKAD-GSEVSFLFYRKYQNQEITDLKQP	274
DB	420	GVKVEVTHRGNMERKRYISGLTAVATARELTFFVDERNRTQKSVVEYFHEFYGRIQHTQLP	479
QY	275	VLYSQPKRRRAGPGITLPGPAMLLPELCYLT--GLTDKWRNDFNMVKDLAVHTRLTPEQRQ	332
DB	480	CLQVGSNR-----PNYLPMEWCKIVGQRYSKRLNERQITALLKV--TCORPIDRE	529
QY	333	REV---GRLLDIYTHKNDNVORELDRDGLSFDSDNLLSPSGRILOTE--KIHQGKRTFDYNP	387
DB	530	KDILQTVQLNDY--AKDNVAQE--FGIKISTSLASVEARILPPWLKYHESGREGTCLP	584
QY	388	QFADWSKETRGAPLISVVKPLDNLWLLIYTRNRYE--AANSILQNLFKVTPPWGMQMR-KAI	444
DB	585	QVGQWNMNK--KMINGCTVNNMICINFSRQVDNLARTFCQELAQMCYVSGMAFNPEPV	642
QY	445	MIEVDORTEAYLRVLQOKVTADT-----QIVVCLLSNRKDKYDAIKKYLCTDCPT	495
DB	643	LPPVSARPEQVEKVLKTRYHDATSKLQSGKEIDLILVILPDNNGSLYGDLLKRICETELGI	702
QY	496	PSOCVWARTLGG-QOYNALATKIALQMNCKMGG-ELWRVD-----IPL---KLWMIVGI	545
DB	703	VSOCCLTKHVFKMSQYMA---NVALKINVKVGRNTVLVDALSRPLVSDRPTIIFGA	759
QY	546	DCYHDMTA--GRRSIAGFVASIN-EGMTRWFSRCIFQDRGOELVDGL-----	589
DB	760	DVTHPEHGSDSPSIAAVVASQDWPEITKYZAGLVCAQAHRQELIQDLFKKWKDPQKGVVT	819
QY	590	KVCLQALRALRAWSCNEYMPSRILIVYRGVGDGQLKTLVNVYVPOQLDCLKSIGGYNPLR	649
DB	820	GEMIKELLIAFRSTGHKPLRIIFYRDGVSEGOYQVLLYELDLAIRKACSLAEGYQPPV	879
QY	650	TVIIVKRVNTRPFAQ-----SGGRQLNPLPGFVIDVEVTRPBYDFPFIQSVARSQSV	703
DB	880	TFVVQQRKHHTLRLEPAQNNDRHSVDRSGNLLPGHVDSKICHPTEFDFVLSHAGIQGTS	939
QY	704	SPTHYNIYDNSGLKPDPHIQELTYKLCHYYNWPGVIRVPAPCOYAHKLAFLVQSIHRE	763
DB	940	RPAHYHVLWDENNFTADGLQSLTNLCVYARCTRSVSIPEPYYAHLAAFEA--RFYME	997
QY	764	PNLSLSNRL 772	
DB	998	PETSDSGSM 1006	
RESULT 15			
DB	ABP64718	standard; Protein; 859 AA.	
XX	AC	ABP64718;	
XX	AC	ABP64718;	
DT	25-FEB-2003	(first entry)	
XX	DE	Human protein SEQ ID 378.	
XX	KW	Human; expressed sequence tag; EST;	
KW	KW	haematopoietic disorder; central nervous system disease; viral infection;	
KW	KW	peripheral nervous system disease; non-healing wound; infectious disease;	
KW	KW	immune deficiency; immune disorder; bacterial infection; allergy; cancer;	
KW	KW	fungal infection; autoimmune disorder; coagulation disorder; neutropenic;	
KW	KW	antiallergic; antiinflammatory; immunosuppressive; neuroprotective;	
KW	KW	cytostatic; haemostatic; virucide; antibacterial; fungicide;	
XX	OS	Homo sapiens.	

GenCore version 5.1.6
Copyright (C) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 6, 2003, 15:52:24 ; Search time 8508 Seconds
(without alignments)
11193.889 Million cell updates/sec

Title: US-10-043-774B-1
Perfect score: 2328
Sequence: 1 agagcttgggtgtgaacac.....accgcctttactacctctaa 2328

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 2888711 seqs, 2045481386 residues

Word size : 0
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_ov.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2328	100.0	2328	9	AF264004 Homo sapi
2	2246	96.5	3421	9	AF387507 Homo sapi
3	2144	92.1	3362	9	AF104260 Homo sapi
4	2144	92.1	3399	9	BC028581 Homo sapi
5	2130	91.5	2579	9	AK093133 Homo sapi
6	2022	9.5	63578	9	AC127071 Homo sapi
7	2022	8.8	193697	2	AC025837 Homo sapi
8	201	8.6	61995	2	AC090565 Homo sapi
9	129	5.5	66611	2	AC090147 Homo sapi
10	125	5.4	66611	2	AC090147 Homo sapi
11	92	4.0	461	6	BD114143 EST and e
12	58	2.5	460	6	AX209866 Sequence
13	35	1.5	3910	10	AB032604 Mus muscu
14	35	1.5	4064	10	AF438405 Mus muscu
15	34	1.5	501	6	AX395819 Sequence
16	32	1.4	39365	2	AC100141 Mus muscu
17	32	1.4	186252	2	AC116715 Mus muscu
18	32	1.4	211319	2	AC111089 Mus muscu
19	32	1.4	259213	2	AC096050 Rattus no
20	26	1.1	39365	2	AC100141 Mus muscu
21	26	1.1	61995	2	AC090565 Homo sapi
22	26	1.1	63578	9	AC127071 Homo sapi
23	26	1.1	193697	2	AC025837 Homo sapi
24	24	1.0	140348	9	AL390725 Human DNA
25	23	1.0	198389	2	AC118198 Mus muscu
26	22	0.9	89501	8	AP006420 Lotus jap
27	22	0.9	135586	9	AC073050 Homo sapi
28	22	0.9	158170	9	AC016757 Homo sapi
29	21	0.9	22	6	AX513354 Sequence
30	21	0.9	1690	3	AK115461 Ciona int
31	21	0.9	9953	1	U67553 Methanococc
32	21	0.9	40907	9	AP000358 Homo sapi
33	21	0.9	50612	9	AL513530 Human DNA
34	21	0.9	81696	9	AF438327_3 Continuation (4 of
35	21	0.9	91748	9	AC004856 Homo sapi
36	21	0.9	101507	9	AP000359 Homo sapi
37	21	0.9	110000	6	AR271569_04 Continuation (5 of
38	21	0.9	110000	9	AF438327_2 Continuation (3 of
39	21	0.9	134058	8	AC037425 Oryza sat
40	21	0.9	159139	2	AC015775 Homo sapi
41	21	0.9	162170	2	AC131452 Strongylo
42	21	0.9	177257	2	AC137121 Mus muscu
43	21	0.9	178976	2	AC118852 Rattus no
44	21	0.9	182770	9	AC104082 Homo sapi
45	21	0.9	185121	2	AC079806 Homo sapi

ALIGNMENTS

RESULT 1
AF264004
LOCUS
DEFINITION Homo sapiens HIWI mRNA, complete cds.
ACCESSION AF264004
VERSION AF264004.1 GI:15216446
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2328)
AUTHORS Sharma,A.K., Nelson,M.C., Brandt,J.E., Wessman,M., Mahmud,N.,
Weller,K.P. and Hoffman,R.
TITLE Human CD34+ stem cells express the hiwi gene, a human homologue of

Db	1561	CAGATGAATCGCAAGATGGAGAGAGCTCTGAGGTTGACATCCCTCGAGTCTGTG	1620		
Qy	1621	ATGATCGTTGGCATCGATTGTTTACCATGACATGACAGCTGGCGGAGGTCAATCGCAGGA	1680		
Db	1621	ATGATCGTTGGCATCGATTGTTTACCATGACATGACAGCTGGCGGAGGTCAATCGCAGGA	1680		
Qy	1681	TTTGTTCGCCAGCATCAATGAAGGATGACCCGCTGGTTCTCAGCTGCGATATTTCCAGGAT	1740		
Db	1681	TTTGTTCGCCAGCATCAATGAAGGATGACCCGCTGGTTCTCAGCTGCGATATTTCCAGGAT	1740		
Qy	1741	AGAGGACAGAGCTGGTAGATGGGCTCAAAAGTCTGCTGCAACGCGCTCTGAGGGCTTGG	1800		
Db	1741	AGAGGACAGAGCTGGTAGATGGGCTCAAAAGTCTGCTGCAACGCGCTCTGAGGGCTTGG	1800		
Qy	1801	AATAGCTGCAATCAGTACATGCCCGCGGATCATCGTGTACCGCGATGGCTAGAGGAC	1860		
Db	1801	AATAGCTGCAATCAGTACATGCCCGCGGATCATCGTGTACCGCGATGGCTAGAGGAC	1860		
Qy	1861	GGCCAGCTGAAAAACACTGTTGAACTACGAAGTGCACAGTTTTTGGATTCTCTAAATCC	1920		
Db	1861	GGCCAGCTGAAACACTGTTGAACTACGAAGTGCACAGTTTTTGGATTCTCTAAATCC	1920		
Qy	1921	ATTGGTAGAGGTTACAAACCTTAGCTAAACGGTAAATTGTTGTGTAAGAAAGAGTGAACACC	1980		
Db	1921	ATTGGTAGAGGTTACAAACCTTAGCTAAACGGTAAATTGTTGTGTAAGAAAGAGTGAACACC	1980		
Qy	1981	AGATTTTTCCTCAGTCTGGAGGAGACTTCAGAAATCCACTTCTCTGGAACAGTATTGAT	2040		
Db	1981	AGATTTTTCCTCAGTCTGGAGGAGACTTCAGAAATCCACTTCTCTGGAACAGTATTGAT	2040		
Qy	2041	GTAGAGGTTACCCAGACCAAGATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAAGT	2100		
Db	2041	GTAGAGGTTACCCAGACCAAGATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAAGT	2100		
Qy	2101	GGTAGTGTCTTCCACACATTAACAATGTCATCTATGACAAACAGCGGCTTGAAGCCAGAC	2160		
Db	2101	GGTAGTGTCTTCCACACATTAACAATGTCATCTATGACAAACAGCGGCTTGAAGCCAGAC	2160		
Qy	2161	CACATACAGCGCTTGACCTCAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTCAATT	2220		
Db	2161	CACATACAGCGCTTGACCTCAAGCTGTGCCACATCTATTACAACTGGCCAGGTGTCAATT	2220		
Qy	2221	CGTGTCTCTGCTCTCTGCGAGTACGCCACCAAGCTGGCTTTTCTTTGTGGCCAGAGTATT	2280		
Db	2221	CGTGTCTCTGCTCTCTGCGAGTACGCCACCAAGCTGGCTTTTCTTTGTGGCCAGAGTATT	2280		
Qy	2281	CACAGAGAGCCAAATCTGTCACTGTCAAAACCGCTTTTACTACCTCTAA	2328		
Db	2281	CACAGAGAGCCAAATCTGTCACTGTCAAAACCGCTTTTACTACCTCTAA	2328		
RESULT 2	AF387507	3421 bp	mRNA	linear	PRI 01-JUL-2001
LOCUS	AF387507				
DEFINITION	Homo sapiens PIWI protein mRNA, complete cds.				
ACCESSION	AF387507				
VERSION	AF387507.1	GI:14579644			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 3421)				
TITLE	Sha, J.H.				
JOURNAL	Cloning and identification of human piwi protein related to testis development				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 3421)				
TITLE	Sha, J.H., Li, J.M. and Zhou, Z.M.				
JOURNAL	Direct Submission				
TITLE	Submitted (30-MAY-2001) Key Lab of Reproductive Medicine, Nanjing Medical University, 140 Han Zhong Road, Nanjing Medical University,				

FEATURES	Location/Qualifiers	Source
1. 3421	/organism="Homo sapiens"	
	/mol_type="mRNA"	
	/db_xref="taxon:9606"	
	/tissue_type="testis"	
	/dev_stage="adult"	
64. 2649	/note="higher expression in adult testis than 6 month old embryo testis"	
	/codon_start=1	
	/product="PIWI protein"	
	/protein_id="AAK69348.1"	
	/db_xref="GI:14579645"	
	/translation="MTGARARARGARGOATQVLVGSTASQOQYVIOPRPOPPAEG ELFRGRTAGTAGTACGQGLISAGFOELSIAREGGRDRPHDLGVNTRONLDAVKE SKTSSGGLIVRLSTNHFRLTSRQWALQYHIDYNPLMEARLRSLALLFOHEDLLIGKC HAFDGLILPKRLQKYVFSKTRNGEDVRITITLITNELPFTTCLQFNILFR LKIMNLOQIGRNYPNDPIDPSHRLVWPGFTTSILOVENSIMCLTDMDQNKSPFK ETVLDFMNFYHTEBEHFQOVSKBLGLVVLTKNNKTVRDVDDIMDQNKSPFK ADGSVFLFYRYKQYNQELTDLKOPVLVSOPKRERGPGTLPGPAMLIPELCYLTGL TDKMENDNVMDLAVHRLTPEQORREVRGLLDYIHKNDNVORELRDWGLSFDNLL SFGRLQTEKHIOGKTFDYNPQADWSKETRGAPLLSVKELDNWLLLYTRENBAAL NSLIQNLKFTVPAGMOMRKAIMTEVDDEAYLRVLOOKVTADTQIVVCLLSNKKO KYDALKYLCYDCTPTPSOCVWARTKQOTVMAIATKALQMNCKGGLVDMRVDPLK LVMIVGIDYHDMTAGRSIAGFVASINEGTRWFRCIFQDRGOELVDGLVQLQAA LEAUNSCNEYMPSRLIVVDGVDGDLKLVNVEVDFDLCLSKIRGVNPRLTAVV KKRNVTRFPAQGGRLQNLPGTVIDVTRPEWDFETVSOAVSGSVSPHTNYIV DNSGKLPHDIQRLTYKLCHYYNWPGVIRVPAPCAHKLAFVVGOSIHRFPNLSLN RLYYL"	
polya signal	3354. .3359	
BASE COUNT	1071 a 661 c 767 g 922 t	
ORIGIN		
Query Match	96.5%; Score 2246; DB 9; Length 3421;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 2296; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
Qy	32	TAGACCATCTTAAAGAAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGTTAAGCACTA 91
Db	353	TAGACCATCTTAAAGAAATCAAAAACAGGTTCTTCAGGCATTATAGTAAGTTAAGCACTA 412
Qy	92	ACCATTTCCGGCTGACATCCCGTCCCGAGTGGGCCCTTATATCAGTATCAGATTGACTATA 151
Db	413	ACCATTTCCGGCTGACATCCCGTCCCGAGTGGGCCCTTATATCAGTATCAGATTGACTATA 472
Qy	152	ACCCACTGATGGAAGCCAGAGACTCCGTTACGCTCTCTTTTCAACAGCAAGATCTAA 211
Db	473	ACCCACTGATGGAAGCCAGAGACTCCGTTACGCTCTCTTTTCAACAGCAAGATCTAA 532
Qy	212	TTGGAAGTGTCACTGCTTTTGATGAACGATATATTTTACCTTAAAGACTACAGCAA 271
Db	533	TTGGAAGTGTCACTGCTTTTGATGAACGATATATTTTACCTTAAAGACTACAGCAA 592
Qy	272	AGGTTACTGAAGTTTATAGTAAGCCCGGAATGGAGAGATGTGAGGATAACGATCACTT 331
Db	593	AGGTTACTGAAGTTTATAGTAAGCCCGGAATGGAGAGATGTGAGGATAACGATCACTT 652
Qy	332	TAACAATGAACCTTCCACCTACATCCACCACTTGTGTTGAGTCTATAATATATTTTCA 391
Db	653	TAACAATGAACCTTCCACCTACATCCACCACTTGTGTTGAGTCTATAATATATTTTCA 712
Qy	392	GGAGGCTTTGAAATCATGAATTTCAACAAATTTGACCAAAATTTATTAACCCCAATG 451
Db	713	GGAGGCTTTGAAATCATGAATTTGACCAAAATTTGACCAAAATTTATTAACCCCAATG 772
Qy	452	ACCAATTTGATATTTCCAAAGTCACAGGTTGTGATTTTGGCTTGGCTTCACTACTTCCATCC 511
Db	773	ACCAATTTGATATTTCCAAAGTCACAGGTTGTGATTTTGGCTTGGCTTCACTACTTCCATCC 832
Qy	512	TTCAAGTATGAAACAGCATCATGCTCTGCACTGACGTTAGCCATTAAGTCTCTGCAAGTG 571

Db 833 TTCAGTATGAAACAGCATCATGCTCTGCACCTGACGCTTAGCCATTAAGTCTCTCGAGTG 892
Qy 572 AGACTGTTTGGATTTTCATGTTTCAACTTTTATCATCAGACAGAAACAATAAATTTCAAG 631
Db 893 AGACTGTTTGGATTTTCATGTTTCAACTTTTATCATCAGACAGAAACAATAAATTTCAAG 952
Qy 632 AACAAAGTTTCCAAAGAACTAAATAGGTTTGTGTTTCTTACCAAGTATATAAATAAGACAT 691
Db 953 AACAAAGTTTCCAAAGAACTAAATAGGTTTGTGTTTCTTACCAAGTATATAAATAAGACAT 1012
Qy 692 ACAGAGTGATGATATGCTGGACGACAGATCCCAAGAGCGACCTTTAAGAAAGCCGAG 751
Db 1013 ACAGAGTGATGATATGCTGGACGACAGATCCCAAGAGCGACCTTTAAGAAAGCCGAG 1072
Qy 752 GCTCTCAAGTCAAGTCTTCTAGAAATACTACAGGAAGCAATACCAACCAAGAGATCACCGACT 811
Db 1073 GCTCTCAAGTCAAGTCTTCTAGAAATACTACAGGAAGCAATACCAACCAAGAGATCACCGACT 1132
Qy 812 TGAAGCAGCCTGTCTTGGTGCAGCCAGCCAGAGAGAGGCGGCGCTTGGGGGACACTGC 871
Db 1133 TGAAGCAGCCTGTCTTGGTGCAGCCAGCCAGAGAGAGGCGGCGCTTGGGGGACACTGC 1192
Qy 872 CAGGCGCTGCCATGCTCATCTGAGCTCTGCTATCTTACAGTCTAACTGATAAATGC 931
Db 1193 CAGGCGCTGCCATGCTCATCTGAGCTCTGCTATCTTACAGTCTAACTGATAAATGC 1252
Qy 932 GTAATGATTTTAAAGTGAAGAGACTTAGCCGCTTATCAACAGACTAACTCCAGAGCAAA 991
Db 1253 GTAATGATTTTAAAGTGAAGAGACTTAGCCGCTTATCAACAGACTAACTCCAGAGCAAA 1312
Qy 992 GGCAGGCTGAAGTGGGAGCGACTCATGATTAATCATTAATAAAGATATGTTCAAGGG 1051
Db 1313 GGCAGGCTGAAGTGGGAGCGACTCATGATTAATCATTAATAAAGATATGTTCAAGGG 1372
Qy 1052 AGCTTCGAGACTGGGTTTGGAGCTTTGATTCCAACTTACTGCTCTTCCAGGAAGATTT 1111
Db 1373 AGCTTCGAGACTGGGTTTGGAGCTTTGATTCCAACTTACTGCTCTTCCAGGAAGATTT 1432
Qy 1112 TGCAAA CAGAAAGATTTCAACAGGTGGAAAAACATTTGATTAACAATCCAAATTTGCAG 1171
Db 1433 TGCAAA CAGAAAGATTTCAACAGGTGGAAAAACATTTGATTAACAATCCAAATTTGCAG 1492
Qy 1172 ATGGTCCAAAGAAACAGAGGTGCACCTAATTAATAGTGTAAAGCACTAGATACTGGC 1231
Db 1493 ATGGTCCAAAGAAACAGAGGTGCACCTAATTAATAGTGTAAAGCACTAGATACTGGC 1552
Qy 1232 TGTGATCTATACGCAAGAAATATGAAGCAGCAATTCATTGATACAAATCTATTTA 1291
Db 1553 TGTGATCTATACGCAAGAAATATGAAGCAGCAATTCATTGATACAAATCTATTTA 1612
Qy 1292 AAGTTACACAGCCATGGGCATCAATGAGAAAGCAATTAATGATGAAGTGGATGACA 1351
Db 1613 AAGTTACACAGCCATGGGCATCAATGAGAAAGCAATTAATGATGAAGTGGATGACA 1672
Qy 1352 GAATGAAGCTACTTAAGATCTTACAGCAAAAGTTCACAGCAGACACCCAGATAGTTC 1411
Db 1673 GAATGAAGCTACTTAAGATCTTACAGCAAAAGTTCACAGCAGACACCCAGATAGTTC 1732
Qy 1412 TCTGCTGTTGTCAAGTAATCGGAAGACAAATACGATGCTATTAAAAATACTGTGTA 1471
Db 1733 TCTGCTGTTGTCAAGTAATCGGAAGACAAATACGATGCTATTAAAAATACTGTGTA 1792
Qy 1472 CAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCGAACCTTAGGCMAACACCAAACTG 1531
Db 1793 CAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCGAACCTTAGGCMAACACCAAACTG 1852
Qy 1532 TCATGGCCATGCTACAAAGATTCCTCCTACAGATGAATCTGAGATGGGAGGAGCTCT 1591
Db 1853 TCATGGCCATGCTACAAAGATTCCTCCTACAGATGAATCTGAGATGGGAGGAGCTCT 1912
Qy 1592 GAGGGTGGACATCCCTCCTGAGCTCTGATGATGCTGTTGGCCTGATGTTACCATGACA 1651
Db 1913 GAGGGTGGACATCCCTCCTGAGCTCTGATGATGCTGTTGGCCTGATGTTACCATGACA 1972

Qy 1652 TGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGATGACCC 1711
Db 1973 TGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGATGACCC 2032
Qy 1712 GCTGGTCTCAGCTGCATATTTTCAAGTATAGAGACAGAGCTGTTAGATGGGCTCAAG 1771
Db 2033 GCTGGTCTCAGCTGCATATTTTCAAGTATAGAGACAGAGCTGTTAGATGGGCTCAAG 2092
Qy 1772 TCTGCTCCAAAGCGGCTCTGAGGGCTTGAATAGTGCATAGTACATGCCCCAGCCGA 1831
Db 2093 TCTGCTCCAAAGCGGCTCTGAGGGCTTGAATAGTGCATAGTACATGCCCCAGCCGA 2152
Qy 1832 TCATGCTTACCGGATGGCGTAGAGACGCGCAGCTGAAACACACTGGTGAACATGAGAG 1891
Db 2153 TCATGCTTACCGGATGGCGTAGAGACGCGCAGCTGAAACACACTGGTGAACATGAGAG 2212
Qy 1892 TGCCACAGTCTTGGATTTGCTTAAATCCATTTGGTAGAGTTTACAACTTAGACTAACGG 1951
Db 2213 TGCCACAGTCTTGGATTTGCTTAAATCCATTTGGTAGAGTTTACAACTTAGACTAACGG 2272
Qy 1952 TAAATGCTGTAAGAAAAAGAGTGAACACAGATTTTGTCTCAGTCTGGAGGAGACTTC 2011
Db 2273 TAAATGCTGTAAGAAAAAGAGTGAACACAGATTTTGTCTCAGTCTGGAGGAGACTTC 2332
Qy 2012 AGAATCCACTCTCTCGAAACAGTATTATGATAGTATGAGTTTACAGACCAAGATGGTATGACT 2071
Db 2333 AGAATCCACTCTCTCGAAACAGTATTATGATAGTATGAGTTTACAGACCAAGATGGTATGACT 2392
Qy 2072 TTTTATCGTGCAGCCAGCTGTGAGAGTGTGAGTGTGTTTCTCCACACATTAACAATGCA 2131
Db 2393 TTTTATCGTGCAGCCAGCTGTGAGAGTGTGAGTGTGTTTCTCCACACATTAACAATGCA 2452
Qy 2132 TCTATGACAAACAGCGGCTGTAAGCCAGACCAATACAGCGTTGACCTTACAGCTGTGCC 2191
Db 2453 TCTATGACAAACAGCGGCTGTAAGCCAGACCAATACAGCGTTGACCTTACAGCTGTGCC 2512
Qy 2192 ACATCTATTAACCTGGCCAGCTGTGAGTGTGTTTCTGCTGCTTCCAGTACGCCACA 2251
Db 2513 ACATCTATTAACCTGGCCAGCTGTGAGTGTGTTTCTGCTGCTTCCAGTACGCCACA 2572
Qy 2252 AGCTGCTTTTCTGCTGTCAGAGTATTACAGAGAGCAATCTGTCACTGTCAAAAC 2311
Db 2573 AGCTGCTTTTCTGCTGTCAGAGTATTACAGAGAGCAATCTGTCACTGTCAAAAC 2632
Qy 2312 GCCTTTACTACCTCTAA 2328
Db 2633 GCCTTTACTACCTCTAA 2649

RESULT 3

AF104260 LOCUS 3362 bp mRNA linear PRI 30-APR-2003
DEFINITION Homo sapiens HIWI mRNA, complete cds.
ACCESSION AF104260
VERSION AF104260.2 GI:18098557
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Cox,D.N., Chao,A., Baker,J., Chang,L., Qiao,D. and Lin,H.
TITLE A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal
JOURNAL Genes Dev. 12 (23), 3715-3727 (1998)
MEDLINE 99059219
PUBMED 9851978
REFERENCE
AUTHORS Qiao,D., Zeeman,A.-M., Deng,W., Looi'jenga,L.H.J. and Lin.H.
TITLE Molecular characterization of hiwi, a human member of the piwi gene family whose overexpression is correlated to seminomas
JOURNAL Oncogene 21 (25), 3988-3999 (2002)

MEDLINE 22033724
PUBMED 12037681
REFERENCE 3 (bases 1 to 3362)
AUTHORS Lin.H.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1998) Cell Biology, Duke University Medical Center, 412 Nanaline Duke Bldg., Research Dr., Durham, NC 27710, USA
REFERENCE 4 (bases 1 to 3362)
AUTHORS Qiao,D., Zeeman,A.-M., Deng,W., Looijenga,L.H.J. and Lin,H.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Cell Biology, Duke University, 412 Nanaline Duke Building, DUMC, Durham, NC 27710, USA
REMARK Nucleotide sequence updated by submitter
COMMENT On Jan 9, 2002 this sequence version replaced gi:4038412.
FEATURES
Location/Qualifiers
1..3362
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q24.33"
/tissue_type="testis"
57..2642
/note="member of the PIWI family"
/codon_start=1
/product="HIWI"
/protein_id="AAC97371.2"
/db_xref="GI:18098558"
/translations="MTGRARARARGARGOETAQLVGSTASQOQGYIQRPQPPAEG
ELPGRGRORAGTAKSOGLOISAGFOELSLAERGRRRDFHDLGVNTRQNDLHVKE
SKTSGSLIVRLSTNHPLNSRQWALVYHIDYNPLMEARRLSALLFOHDLIGKC
HAFDGLFLPKLQKYVPSKTRNGEDVRITLTNLPSTPCQFVNIIFRR
LLKIMLQQIGRNYINPDIDIPSHRLVWPFTTSILOENSIIMCLDVSXKVRLS
ETVLDFMFNYHTEHKEFQESKELIGLIVLTKNIKYRVDIDWDONPKSTFK
ADGGSVFLVYKQVNOEITDLKQFVLVSPQKRRPGGTLFPGPMLIPELCYLGL
TDKRNDFNVKMLAVNHEITLPORQREVGLIDYHKNDNVQRELRLDWGLSPDSML
SFSGRILQTEKHOGKTFDYNQPADMSKETRGAPLISVKPLDMLLIIVTRNVEAA
NSILQNLKTPPANGQMRVAMLEVDDEPEAVLRVLQKVADTQIVLCLSSNKD
KIDAIKYLCTDPTQCVQVARTLQKQVMAIATKIALQOMKMGGEUWRVDIELK
LVNIVGDYHDMTAGRSIAGFVAGSINEMTFRSRIFQDRGELVLDGLKVCLOAA
LRAWNSNEYMSRRIIVYRDGVGDGLKTVNTEVDFQDLKSIORGNPRLTIVV
KRVNTRFFAQSGRLQNPGLTVIDVEVTRPEWDFIVSQVRSVSPHTNVIY
DNGLKPDHTQLRYLKLHIYNWPGVIRVAPQYAHKLAFLVGSIHREPNSLSN
RLYL"

BASE COUNT 1026 a 659 c 759 g 918 t

ORIGIN
Query Match 92.1%; Score 2144; DB 9; Length 3362;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2294; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 32 TAGACCATGTTAAAGATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTA 91
Db 346 TAGACCATGTTAAAGATCAAAAACAGGTTCTTCAGGCATTATAGTAAGGTTAAGCACTA 405

Qy 92 ACCATTTCCGGCTGACATCCCGTCCCGAGTGGCGCTTATATCAGTATCAGATCACTATA 151
Db 406 ACCATTTCCGGCTGACATCCCGTCCCGAGTGGCGCTTATATCAGTATCAGATCACTATA 465

Qy 152 ACCCACTGATGGAAGCCAGAGAGATCCCGTTCAGGCTCTCTTTTCAACACAGATCTAA 211
Db 466 ACCCACTGATGGAAGCCAGAGATCCCGTTCAGGCTCTCTTTTCAACACAGATCTAA 525

Qy 212 TTGGAAGTGTGATGCTTTTGTAGTGAACGATATATTTTACCTAAAGACTACAGCAA 271
Db 526 TTGGAAGTGTGATGCTTTTGTAGTGAACGATATATTTTACCTAAAGACTACAGCAA 585

Qy 272 AGGTTACTGAAGTTTGTAGTGAACCGGAAATGGAGAGATGTGAGGATACGATCACTT 331
Db 586 AGGTTACTGAAGTTTGTAGTGAACCGGAAATGGAGAGATGTGAGGATACGATCACTT 645

Qy 332 TAACAAATGAACCTCCACCTACATCAACCACTTGTTCAGCTTCTATAATATTTTCA 391

Db 646 TAACAAATGAACCTCCACCTACATCAACCACTTGTTCAGCTTCTATAATATTTTCA 705

Qy 392 GGAGGCTTTGAAATCATGAATTTGCACAAATTTGGACGAATATTTAAACCCAAATG 451

Db 706 GGAGGCTTTTGAATCATGAATTTGCACAAATTTGGACGAATATTTAAACCCAAATG 765

Qy 452 ACCCAATGATATTCCAAGTCACAGTTGGCTGATTTGGCTGGCTTCACTACTTCATCC 511

Db 766 ACCCAATGATATTCCAAGTCACAGTTGGCTGATTTGGCTGGCTTCACTACTTCATCC 825

Qy 512 TTCAAGTATGAAACAGCATCATGCTCTGCATGAGCTTACAGCATAAAGTCCCTTCGAAGTG 571

Db 826 TTCAAGTATGAAACAGCATCATGCTCTGCATGAGCTTACAGCATAAAGTCCCTTCGAAGTG 885

Qy 572 AGACTGTTTGGATTTCACTTTTATCATCAGACAGAAACATAAATTTCAAG 631

Db 886 AGACTGTTTGGATTTCACTTTTATCATCAGACAGAAACATAAATTTCAAG 945

Qy 632 AACCAAGTTTCCAAAGAACTAATAGTTTGTCTTACCAAGTATAACAATAAGACAT 691

Db 946 AACCAAGTTTCCAAAGAACTAATAGTTTGTCTTACCAAGTATAACAATAAGACAT 1005

Qy 692 ACAGAGTGGATGATATTTGATCTGGACCCAGAAATCCCAAGAGCCCTTTAAGAAAGCCGAG 751

Db 1006 ACAGAGTGGATGATATTTGATCTGGACCCAGAAATCCCAAGAGCCCTTTAAGAAAGCCGAG 1065

Qy 752 GCTCTGAAGTCAGCTCTTAGATATCTACAGAGCAATACCAAGAGATCAACCGACT 811

Db 1066 GCTCTGGGTGAGCTCTTTAGATATCTACAGAGCAATACCAAGAGATCAACCGACT 1125

Qy 812 TGAAGCAGAGCTGCTTGTGTCAGCAGCCCAAGAGAGCGGGCCCTGGGGGACACTGC 871

Db 1126 TGAAGCAGAGCTGCTTGTGTCAGCAGCCCAAGAGAGCGGGCCCTGGGGGACACTGC 1185

Qy 872 CAGGGCTCCCATGCTCAITTCCTGAGCTCTGCTATCTTACAGGCTTAATGATAAATGC 931

Db 1186 CAGGGCTCCCATGCTCAITTCCTGAGCTCTGCTATCTTACAGGCTTAATGATAAATGC 1245

Qy 932 GTATGATTTAAGTCGATGAAGAGCTTACCGCTTATCAACAGACTTAATCCAGAGCAA 991

Db 1246 GTATGATTTAAGTCGATGAAGAGCTTACCGCTTATCAACAGACTTAATCCAGAGCAA 1305

Qy 992 GGCAGCGTGAAGTGGGACGACTCAITGATATCAATTAATAAAGCAATATGTTCAAAGGG 1051

Db 1306 GGCAGCGTGAAGTGGGACGACTCAITGATATCAATTAATAAAGCAATATGTTCAAAGGG 1365

Qy 1052 AGTTTCGAGACTGGGGTTTGAAGCTTTGATTTCCAACTTATCTGCTCTCTCAGGAAGATTT 1111

Db 1366 AGTTTCGAGACTGGGGTTTGAAGCTTTGATTTCCAACTTATCTGCTCTCTCAGGAAGATTT 1425

Qy 1112 TGCAACAGAGAAAGATTCACAGGTGGAAACATTTGATTAATCAATCCCAATTTTCAG 1171

Db 1426 TGCAACAGAGAAAGATTCACAGGTGGAAACATTTGATTAATCAATCCCAATTTTCAG 1485

Qy 1172 ATTGGTCCAAAGAAACAGAGGTGCACCAATTAATAGTGTTAAGCCACTAGATACTGGC 1231

Db 1486 ATTGGTCCAAAGAAACAGAGGTGCACCAATTAATAGTGTTAAGCCACTAGATACTGGC 1545

Qy 1232 TGTGTATCTATACCGGAAGAAATTTGAAGCAGCCCAATTCATTTGATACAAAATCTATTTA 1291

Db 1546 TGTGTATCTATACCGGAAGAAATTTGAAGCAGCCCAATTCATTTGATACAAAATCTATTTA 1605

Qy 1292 AGTTTACACAGCCATGGGCATGCAATGAGAAAGCAATTAATGATTAAGTGGATGACA 1351

Db 1606 AGTTTACACAGCCATGGGCATGCAATGAGAAAGCAATTAATGATTAAGTGGATGACA 1665

Qy 1352 GAACAGGCTTACTTTAAAGTCTTACAGCAAAGGTTCACAGCAGACACCCAGATAGTTG 1411

Db 1666 GAACAGGCTTACTTTAAAGTCTTACAGCAAAGGTTCACAGCAGACACCCAGATAGTTG 1725

Qy 1412 TCTGTCTGTTGTCAAGTAATTCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTA 1471

Db 423 ACCATTTCGGCTGACATCCCGTCCCGAGTGGCGCTTATATCAGTATCATCAATGACTATA 482
Qy 152 ACCACGTGATGGAAGCCAGAGACTCGGTTCAAGTCTCTCTTTTCAACAGAGATCTAA 211
Db 483 ACCACGTGATGGAAGCCAGAGACTCGGTTCAAGTCTCTCTTTTCAACAGAGATCTAA 542
Qy 212 TTGGAAAGTGTCTATGCTTTTGTGGAAGAGATATTTATTTTACCTTAAAGAGACTACAGCAA 271
Db 543 TTGGAAAGTGTCTATGCTTTTGTGGAAGAGATATTTATTTTACCTTAAAGAGACTACAGCAA 602
Qy 272 AGTTACTGAAGTTTATGTAAGACCCGGAATGGAGAGAGTGGAGATTAACGATCACTT 331
Db 603 AGTTACTGAAGTTTATGTAAGACCCGGAATGGAGAGAGTGGAGATTAACGATCACTT 662
Qy 332 TACAAATGAATTCACCTACATCAACCACTTTGTTGCAAGTCTTAAATATTTTCA 391
Db 663 TACAAATGAATTCACCTACATCAACCACTTTGTTGCAAGTCTTAAATATTTTCA 722
Qy 392 GGAGGCTTTTGAATATCATGAATTTGCAACAAATTTGACGAAATTTATTAACCCAAATG 451
Db 723 GGAGGCTTTTGAATATCATGAATTTGCAACAAATTTGACGAAATTTATTAACCCAAATG 782
Qy 452 ACCCAATGATATTCGAAGTCAAGTGTGTAATTTGGCTTGGCTTCACTACTTCCATCC 511
Db 783 ACCCAATGATATTCGAAGTCAAGTGTGTAATTTGGCTTGGCTTCACTACTTCCATCC 842
Qy 512 TTCAGTATGAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCTTTCGAGTG 571
Db 843 TTCAGTATGAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCTTTCGAGTG 902
Qy 572 AGACTGTTTGGATTTTCATGTTTCACTTTTATCATCAGAGAGAACATATAATTTCAAG 631
Db 903 AGACTGTTTGGATTTTCATGTTTCACTTTTATCATCAGAGAGAACATATAATTTCAAG 962
Qy 632 AACAGTTTCCAAAGACTAATAGTTAGTTAGTTTCTTACCAAGTATTAACATTAAGACAT 691
Db 963 AACAGTTTCCAAAGACTAATAGTTAGTTAGTTTCTTACCAAGTATTAACATTAAGACAT 1022
Qy 692 ACAGAGTGAATGATTTGACTGGAGCCAGATCCCAAGAGACCTTTAAGAAAGCCGAG 751
Db 1023 ACAGAGTGAATGATTTGACTGGAGCCAGATCCCAAGAGACCTTTAAGAAAGCCGAG 1082
Qy 752 GCTCTGAAGTCACTTCTTGAATTAAGTACTACAGAGAGCAATAACAACAGAGATCAACGACT 811
Db 1083 GCTCTGAAGTCACTTCTTGAATTAAGTACTACAGAGAGCAATAACAACAGAGATCAACGACT 1142
Qy 812 TGAAGCAGCTGTCTTGTGTCAGCAGCCCAAGAGAGCGCGGCGCTTGGGGGACACTGC 871
Db 1143 TGAAGCAGCTGTCTTGTGTCAGCAGCCCAAGAGAGCGCGGCGCTTGGGGGACACTGC 1202
Qy 872 CAGGCGCTGCCATGCTCATCTGAGCTCTGCTATCTTACAGTCTTAACTGATTAAGTGC 931
Db 1203 CAGGCGCTGCCATGCTCATCTGAGCTCTGCTATCTTACAGTCTTAACTGATTAAGTGC 1262
Qy 932 GTAATGATTTTAAAGTGAAGAGACTTAGCGGTTTACACAGACTTAACTCCAGAGCAA 991
Db 1263 GTAATGATTTTAAAGTGAAGAGACTTAGCGGTTTACACAGACTTAACTCCAGAGCAA 1322
Qy 992 GGCAGCGTGAAGTGGGAGACTCATTTGATTAATTTATTAATAAAGAGATATTTCAAGGG 1051
Db 1323 GGCAGCGTGAAGTGGGAGACTCATTTGATTAATTTATTAATAAAGAGATATTTCAAGGG 1382
Qy 1052 AGCTTCGAGACTGGGTTTGAAGTGTGATTTTCAAGTCTTCTTCTCAGAGAGATTT 1111
Db 1383 AGCTTCGAGACTGGGTTTGAAGTGTGATTTTCAAGTCTTCTTCTCAGAGAGATTT 1442
Qy 1112 TGCACACAGAAAGATTCACCAAGTGGAAAGCAATTTGATTACAAATCCAAATTTGAG 1171
Db 1443 TGCACACAGAAAGATTCACCAAGTGGAAAGCAATTTGATTACAAATCCAAATTTGAG 1502
Qy 1172 ATTGGTCCAAAGAAACAAGAGTGCACCAATTAATTTAGTTTAAAGCCACTAGATACTGC 1231
Db 1503 ATTGGTCCAAAGAAACAAGAGTGCACCAATTAATTTAGTTTAAATCCCACTAGATACTGC 1562

Qy 1232 TGTGATCTATACGGAAGAAATTTATGAAGCAGCCAAATTCATTGATCAAAAATCTATTTA 1291
Db 1563 TGTGATCTATACGGAAGAAATTTATGAAGCAGCCAAATTCATTGATCAAAAATCTATTTA 1622
Qy 1292 AAGTTACACAGCCATGGGCATGCAAAATGAGAAAAGCAATAATGATGAAGTGGATGACA 1351
Db 1623 AAGTTACACAGCCATGGGCATGCAAAATGAGAAAAGCAATAATGATGAAGTGGATGACA 1682
Qy 1352 GAATGAAGCTTACTTAAGAGTCTTACAGCAAAAGGTACAGCAGACACCCAGATAGTTG 1411
Db 1683 GAATGAAGCTTACTTAAGAGTCTTACAGCAAAAGGTACAGCAGACACCCAGATAGTTG 1742
Qy 1412 TCTGCTGTGTGTCAGAGTAAATCGAAGGCAAAATACCATGCTATTTAAAAATACTGTGTA 1471
Db 1743 TCTGCTGTGTGTCAGAGTAAATCGAAGGCAAAATACCATGCTATTTAAAAATACTGTGTA 1802
Qy 1472 CAGATTGCCCTACCCCAAGTCAAGTGTGTGGTGGCCGGAACCTTAGGCAACACGAACTG 1531
Db 1803 CAGATTGCCCTACCCCAAGTCAAGTGTGTGGTGGCCGGAACCTTAGGCAACACGAACTG 1862
Qy 1532 TCAATGCCATTTGCTACAAAGATTTCCCTACAGATGAACCTGCARAGATGGGAGGAGCTCT 1591
Db 1863 TCAATGCCATTTGCTACAAAGATTTCCCTACAGATGAACCTGCARAGATGGGAGGAGCTCT 1922
Qy 1592 GGAGGTTGACATCCCTTGAAGTCTGATGATCGTTGGCATCGATTGTTACCATGACA 1651
Db 1923 GGAGGTTGACATCCCTTGAAGTCTGATGATCGTTGGCATCGATTGTTACCATGACA 1982
Qy 1652 TGACAGCTGGCGGAGGTCAATCGAGGATTTGTTGCCAGCATCAATGAAGGAGTGAACC 1711
Db 1983 TGACAGCTGGCGGAGGTCAATCGAGGATTTGTTGCCAGCATCAATGAAGGAGTGAACC 2042
Qy 1712 GCTGTTCTCACGCTGCATATTTACAGATAGAGACAGGAGCTGTAGATGGGCTCAAAG 1771
Db 2043 GCTGTTCTCACGCTGCATATTTACAGATAGAGACAGGAGCTGTAGATGGGCTCAAAG 2102
Qy 1772 TCTGCTGCAAGCGGCTCTGAGGCTTGGAAATAGCTGCAATGAGTACATGCCAGCCGGA 1831
Db 2103 TCTGCTGCAAGCGGCTCTGAGGCTTGGAAATAGCTGCAATGAGTACATGCCAGCCGGA 2162
Qy 1832 TCATCTGTATACCCGATGGCGTAGGAGACGGCCAGCTGAAAAACACTGGTGAACCTACGAAG 1891
Db 2163 TCATCTGTATACCCGATGGCGTAGGAGACGGCCAGCTGAAAAACACTGGTGAACCTACGAAG 2222
Qy 1892 TGCCACAGTTTTTGGATTTGTTAAATTCATTTGGTAGAGGTTACAAACCTTAGACTACGG 1951
Db 2223 TGCCACAGTTTTTGGATTTGTTAAATTCATTTGGTAGAGGTTACAAACCTTAGACTACGG 2282
Qy 1952 TAATTTGGTGAAGAAAGAGTGAACACCAAGATTTTTTGTCTGCTGAGGAGAGACTTC 2011
Db 2283 TAATTTGGTGAAGAAAGAGTGAACACCAAGATTTTTTGTCTGCTGAGGAGAGACTTC 2342
Qy 2012 AGAATCCATTCCTCGAACAGTTTATGATGTAGAGTTTACAGACCAAGATTTGGTATGACT 2071
Db 2343 AGAATCCATTCCTCGAACAGTTTATGATGTAGAGTTTACAGACCAAGATTTGGTATGACT 2402
Qy 2072 TTTTATTCGTGAGCCAGGCTGTGAGAGTGTGTTTCTCCACACATTAAGATGCA 2131
Db 2403 TTTTATTCGTGAGCCAGGCTGTGAGAGTGTGTTTCTCCACACATTAAGATGCA 2462
Qy 2132 TCTATGACACAGCGGCTTGAAGCCAGACCAATACAGCGCTTGACCTCAAGCTGTGCC 2191
Db 2463 TCTATGACACAGCGGCTTGAAGCCAGACCAATACAGCGCTTGACCTCAAGCTGTGCC 2522
Qy 2192 ACATCTATTAACATGGCCAGGTTGCTATTCGTGTTCTCTGCTCCCTGAGTACGCCCAACA 2251
Db 2523 ACATCTATTAACATGGCCAGGTTGCTATTCGTGTTCTCTGCTCCCTGAGTACGCCCAACA 2582
Qy 2252 AGCTGGCTTTCTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCCTACTGTCAAAACC 2311
Db 2583 AGCTGGCTTTCTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCCTACTGTCAAAACC 2642


```

QY 2312 GCCTTTACTACTCTTAA 2328
Db 2643 GCCTTTACTACTCTTAA 2659

RESULT 5
AK093133
LOCUS Homo sapiens cDNA FLJ35814 fis, clone TEST12006083, highly similar
to Mus musculus mRNA for MIWI (piwi).
ACCESSION AK093133
VERSION AK093133.1 GI:21751907
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,
Arita, M., Musashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R.,
Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,
Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,
Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B.,
Suzuki, Y., Sugano, S., Nagahari, K., Masuno, Y., Negai, K. and
Isogai, T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2579)
Isogai, T. and Yamamoto, J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(S-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
Location/Qualifiers
1 . 2579
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TEST12006083"
/tissue_type="testis"
/clone_lib="TEST12"
/notes="cloning vector: pME18SFL3"
54..2543
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAC04068.1"
/db_xref="GI:21751908"
/translation="MTGRARARARGRQETAOLVGSTASQDPGYIQRPQPPARG
ELFGRGQRGTATKASQGLQISAGFQELSLAERGRDRFDHLYGNTRQNLHVKC
SKTSGSITVRLSTHPRLTSPOMALYQYHIDYINPLMEARLRSLFOHEDLIGKC
HAPGTLILFLPRLQOKATEVSKTRNGEDVRIITLTNRLPPTSPCLQFYNIIPR
LLKTMNLOQIGRNYNDPDIIPSHRLIWPGETTSLIOYENSIMLCTDPSHKVLES
ETVLDPMFNFTQTEHFKQVSKELIGLVLLTKYNNKRYRVDNDINDONPKSTFKK
ADGSEVFLFYKQYNOEITDLQPVLSQPKRRPGPGLPFPAMLIPELCYLTGL
TDKMRNDFNMWKLAVETRLTPEORVEGLIDYIHKNDNVQRELSDWGLSFDNL
SFSGRILQTEKHOGKTFDYNPFDVDSKETGAPLISVKPLDNLWLLIYTRNRYEA
NSLIQNLKPTPMGMOMRKAI MLEVDDRTEAYLRVLQKQADTQIIVCLLSNRKD
KYDALKYLCTDCTPSCQCVVARTLGKQQTWALATKIALQMNCKMGSELWRVDIPLK
LVMTVGIDCYDHMTAGRSAGFVASINEGMTWFSRCIFQDRQELVDGLVCLQLA
LRWNSCNEYMPSRIIIVRDGQGLKTLVNYEVPQFLDCLKISIGRGNPRLTVIV
KKRVNTRFFAOSGRLQNPILPGTIVDVEVTRPEWDFIVSQAVERSGVSPTHNYIV
DNSGLKPDHQLRIKLCHIYNNPVSASTC"

BASE COUNT 797 a 542 c 613 g 627 t
ORIGIN

```

```

Query Match 91.5%; Score 2130; DB 9; Length 2579;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2180; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 32 TAGACCATGTTAAAGAAATCAAAAAAGGTTCTTCAGGCATTATAGTAAGGTTAAAGCACTA 91
Db 343 TAGACCATGTTAAAGAAATCAAAAAAGGTTCTTCAGGCATTATAGTAAGGTTAAAGCACTA 402
QY 92 ACCATTTCGGCTGCATCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 151
Db 403 ACCATTTCGGCTGCATCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 462
QY 152 ACCACTGATGAGGAGCAGAGACATCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 211
Db 463 ACCACTGATGAGGAGCAGAGACATCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGTCCCGT 522
QY 212 TTGGAAAGTGTCATGCTTTTGTGATGGAACGATATATTTTACCTAAAGACTACAGCAAA 271
Db 523 TTGGAAAGTGTCATGCTTTTGTGATGGAACGATATATTTTACCTAAAGACTACAGCAAA 582
QY 272 AGGTTACTGAAGTTTTTATAGAACCCCGGAATGGAGAGGATGTGAGGATAACGATCACTT 331
Db 583 AGGTTACTGAAGTTTTTATAGAACCCCGGAATGGAGAGGATGTGAGGATAACGATCACTT 642
QY 332 TAACAAATGAATTCACCTACATCAACCACTTGTTCAGTCTCTATATATTTTCA 391
Db 643 TAACAAATGAATTCACCTACATCAACCACTTGTTCAGTCTCTATATATTTTCA 702
QY 392 GGAGGCTTTTGAATAATCATGAATTTGCAACAAATGGAGCAAAATATTTAAACCCAAATG 451
Db 703 GGAGGCTTTTGAATAATCATGAATTTGCAACAAATGGAGCAAAATATTTAAACCCAAATG 762
QY 452 ACCCAATGATATTCCAAGTCAAGGTTGGTGAATTTGGCTTGCCTTCACTTCCATCC 511
Db 763 ACCCAATGATATTCCAAGTCAAGGTTGGTGAATTTGGCTTGCCTTCACTTCCATCC 822
QY 512 TTCAGTATGAAACAGACATCTCTGCACTGACCTTAGCCCTAAGTCCCTTCGAAGTG 571
Db 823 TTCAGTATGAAACAGACATCTCTGCACTGACCTTAGCCCTAAGTCCCTTCGAAGTG 882
QY 572 AGACTCTTTTGGATTTTCATGTTCAACTTTTATCATCAGACAGAAACATAAATTTCAAG 631
Db 883 AGACTCTTTTGGATTTTCATGTTCAACTTTTATCATCAGACAGAAACATAAATTTCAAG 942
QY 632 AACCAATGTTCCAAAGAACTAATAGTTTGTGTTCTTACCAGTATATAAATAAGACAT 691
Db 943 AACCAATGTTCCAAAGAACTAATAGTTTGTGTTCTTACCAGTATATAAATAAGACAT 1002
QY 692 ACAGAGTGGATGATATTTGCTGGGACCAAGTCCCAAGAGCACCTTTAAGAAAGCCGAGG 751
Db 1003 ACAGAGTGGATGATATTTGCTGGGACCAAGTCCCAAGAGCACCTTTAAGAAAGCCGAGG 1062
QY 752 GCTCTGAAGTCAGCTTTCTTAGAATACTACAGGAAGCAATAACAACAGAGATCACCGACT 811
Db 1063 GCTCTGAAGTCAGCTTTCTTAGAATACTACAGGAAGCAATAACAACAGAGATCACCGACT 1122
QY 812 TGAAGCAGCGCTGTTTGGTTCAGCCAGCCAGAGAGGCGGGCCCTGGGGGACACTGC 871
Db 1123 TGAAGCAGCGCTGTTTGGTTCAGCCAGCCAGAGAGGCGGGCCCTGGGGGACACTGC 1182
QY 872 CAGGGCTCGCATGCTCATTCTCGCTGCTCTCTTCTTACAGTCTAACTGATAAATGC 931
Db 1183 CAGGGCTCGCATGCTCATTCTCGCTGCTCTCTTCTTACAGTCTAACTGATAAATGC 1242
QY 932 GTAATGATTTTAAAGTGATGAAGACTTAGCCCTGCTCATACAGACTAACTCCAGAGCAAA 991
Db 1243 GTAATGATTTTAAAGTGATGAAGACTTAGCCCTGCTCATACAGACTAACTCCAGAGCAAA 1302
QY 992 GGCAGCGTGAGGTGGGACGACCTATTGATTACATTAATAAAGCGATAATGTTCAAAGGG 1051
Db 1303 GGCAGCGTGAGGTGGGACGACCTATTGATTACATTAATAAAGCGATAATGTTCAAAGGG 1362

```

QY 1052 AGCTTCGAGCTGGGTTTGAGCTTTGATTCACAACTTACTGTCTCTCTCAGGAAGATTT 1111
DB 1363 AGCTTCGAGCTGGGTTTGAGCTTTGATTCACAACTTACTGTCTCTCAGGAAGATTT 1422
QY 1112 TGCAACACAGAAAAGATTACCAAGGTGGAAAAACATTTGATTACATTCACATTTGCGAG 1171
DB 1423 TGCAACACAGAAAAGATTACCAAGGTGGAAAAACATTTGATTACATTCACATTTGCGAG 1482
QY 1172 ATTTGGTCCAAAGAAAACAGAGGTGCACCAATTAATTAGTTAGTCCCACTAGATACTGGC 1231
DB 1483 ATTTGGTCCAAAGAAAACAGAGGTGCACCAATTAATTAGTTAGTCCCACTAGATACTGGC 1542
QY 1232 TGTGTGCTATATACCGGAAGAAAATATGACGACGCAATTCATTGATGACAAAATCTATTTA 1291
DB 1543 TGTGTGCTATATACCGGAAGAAAATATGACGACGCAATTCATTGATGACAAAATCTATTTA 1502
QY 1292 AAGTTACACAGCCATGGCGATCAATGAGAAAAGCAATTAATGATGAAGTGGATGACA 1351
DB 1603 AAGTTACACAGCCATGGCGATCAATGAGAAAAGCAATTAATGATGAAGTGGATGACA 1662
QY 1352 GAATGAGCTTACTTAAGAGTCTTACGACAAAAGTTCACAGCAGACACCCAGATAGTTG 1411
DB 1663 GAATGAGCTTACTTAAGAGTCTTACGACAAAAGTTCACAGCAGACACCCAGATAGTTG 1722
QY 1412 TCTGTCTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTA 1471
DB 1723 TCTGTCTGTTGTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTA 1782
QY 1472 CAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCGAACTTAGGCAACACCAAACTG 1531
DB 1783 CAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCGAACTTAGGCAACACCAAACTG 1842
QY 1532 TCATGGCCATGCTACAAAGATTGCCCTACAGATGAATCGAGATGGGAGAGAGTCT 1591
DB 1843 TCATGGCCATGCTACAAAGATTGCCCTACAGATGAATCGAGATGGGAGAGAGTCT 1902
QY 1592 GGAGGTGGACATCCCTCGAGCTCGTGTGATGCTGTGGTGCATCGATTGTACCATGACA 1651
DB 1903 GGAGGTGGACATCCCTCGAGCTCGTGTGATGCTGTGGTGCATCGATTGTACCATGACA 1962
QY 1652 TGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCC 1711
DB 1963 TGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGATGACCC 2022
QY 1712 GCTGTTTCTCAGCTGCTATATTCAGATAGAGACAGAGAGCTGGTATAGTGGCTCAAG 1771
DB 2023 GCTGTTTCTCAGCTGCTATATTCAGATAGAGACAGAGAGCTGGTATAGTGGCTCAAG 2082
QY 1772 TCTGCTGCAAGCGGCTCTGAGGCTTGGATAGCTGCAATGAGTACATGCCCCAGCCGGA 1831
DB 2083 TCTGCTGCAAGCGGCTCTGAGGCTTGGATAGCTGCAATGAGTACATGCCCCAGCCGGA 2142
QY 1832 TCATCGTGTACCGGATGGCGTAGGACGCGCAGCTGAAACACTGGTGAACCTACGAAG 1891
DB 2143 TCATCGTGTACCGGATGGCGTAGGACGCGCAGCTGAAACACTGGTGAACCTACGAAG 2202
QY 1892 TGCCACAGTTTTTGGATGTTCTAAATCCATTTGTTAGAGTTTCAACCCCTAGCTAAGG 1951
DB 2203 TGCCACAGTTTTTGGATGTTCTAAATCCATTTGTTAGAGTTTCAACCCCTAGCTAAGG 2262
QY 1952 TAATTTGGTGAAGAAAAGAGTGAACACAGATTTTTTGTCTGCTGGAGGAAGCTTC 2011
DB 2263 TAATTTGGTGAAGAAAAGAGTGAACACAGATTTTTTGTCTGCTGGAGGAAGCTTC 2322
QY 2012 AGAATCCACTTCTCGGAACAGTTATTTGATGTAGAGTTTACAGACCCAGAAATGGATGACT 2071
DB 2323 AGAATCCACTTCTCGGAACAGTTATTTGATGTAGAGTTTACAGACCCAGAAATGGATGACT 2382
QY 2072 TTTTATCGTGAGCAGCTGTGAGAGTGGTATGTTTCTCCACACATTAACATGTC 2131
DB 2383 TTTTATCGTGAGCAGCTGTGAGAGTGGTATGTTTCTCCACACATTAACATGTC 2442
QY 2132 TCTATGACAAACGCGGCTTGAGCCAGACCATACACGCGCTTGACCTACAGCTGTGTC 2191

DB 2443 TCTATGACAAACGCGGCTTGAGCCAGACCATACAGCGCTTGACCTACAGCTGTGTC 2502
QY 2192 ACATCTATTACAACTGGCCAG 2212
DB 2503 ACATCTATTACAACTGGCCAG 2523

RESULT 6

AC127071 63578 bp DNA linear PRI 03-SEP-2002
Homo sapiens 12 BAC RP11-200K12 (Roswell Park Cancer Institute
Human BAC Library) complete sequence.

AC127071
AC127071.3 GI:22657480
HTG.
Homo sapiens (human)
Homo sapiens
Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 63578)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaraturge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbieri,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyie,M.D., Dathorne,S.R., David,R.,
David,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douchwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,
Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,
Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,F., Garza,N.,
Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Haylak,P.,
Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M.,
Holloway,C., Hollins,I., Jackson,L.E., Jacobson,B., Jia,Y.,
Hume,J., Ioshikhes,I., Johnson,B., Honsi,F., Howard,S., Huber,J., Hulyk,S.,
Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,
King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N.,
Lieu,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O.,
Lien,C., Liu,J., Liu,W., Lousheed,H., Lozada,R.J., Lu,X.,
Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P.,
Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,
Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S.,
Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,
Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,
Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okwuonu,G.,
Ogunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Fu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shim,C., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Umani,K., Vasquez,L., Vera,V., Villalobos,B., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kuchelapati,R.,
Wainstock,G., and Gibbs,R.

TITLE

Unpublished

2 (bases 1 to 63578)

Worley,K.C.

Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department

of Molecular and Human Genetics, Baylor College of Medicine, One

Baylour Plaza, Houston, TX 77030, USA

3 (bases 1 to 63578)

Worley,K.C.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (01-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 63578)
Worley, K.C.

Direct Submission
Submitted (03-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Sep 3, 2002 this sequence version replaced gi:21908383.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
of STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES
source
Location/Qualifiers
1. .63578
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-200K12"
/complement(1. .2013)
/note="overlaps bases 1. .2013 of clone AC026764"
/functions="clone overlap"
1694. .1985
/standard_name="G64712"
15975. .16062
/function="single clone coverage"
19548. .19627
/note="sized by PCR to be 80 bps and confirmed by restriction digest"
/function="unresolved dinucleotide repeat"
50966. .51098
/standard_name="RH102272"
complement(61575. .63578)
/note="overlaps bases 190771. .192774 of clone AC063926"
/function="clone overlap"
61711. .62052
/standard_name="RH122074"

BASE COUNT 16941 a 13136 c 14160 g 19341 t
ORIGIN

Query Match 9.5%; Score 222; DB 9; Length 63578;
Best Local Similarity 100.0%; Pred. No. 5.6e-121;
Matches 222; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 CAGGTTCTTCAGGCATTATAGTAAAGTAAAGCACTAAACCATTCGCGGTGACATCCCGTC 115
|||
Db 25305 CAGGTTCTTCAGGCATTATAGTAAAGTAAAGCACTAAACCATTCGCGGTGACATCCCGTC 25364
|||

QY 116 CCCAGTGGCCCTTATATCATGATATACATTCAGTATACCCACTGATGGAGCCAGAGAC 175
|||
Db 25365 CCCAGTGGCCCTTATATCATGATATACATTCAGTATACCCACTGATGGAGCCAGAGAC 25424
|||

QY 176 TCGTTTCAGCTCTCTCTTTTCAACAGAGATCTAATTCGAAAGTGTATGCTTTTGATG 235
|||
Db 25425 TCGTTTCAGCTCTCTTTTCAACAGAGATCTAATTCGAAAGTGTATGCTTTTGATG 25484
|||

QY 236 GAACGATATATTTTACCTAAAGACTACAGCAAAAGGTTA 277
|||
Db 25485 GAACGATATATTTTACCTAAAGACTACAGCAAAAGGTTA 25526
|||

RESULT 7
AC025837 193697 bp DNA linear HTG 06-MAY-2001
LOCUS Homo sapiens clone RP11-200K12, WORKING DRAFT SEQUENCE, 29
DEFINITION unordered pieces.
AC025837
ACCESSION AC025837.3 GI:8072609
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 193697)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome, clone RP11-200K12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 193697)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Bouckhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehocsky, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marguis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (16-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced gi:7342026.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>

Contact: sequence submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5437
Center clone name: 200_K12

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 174594 bases at least Q40

Consensus quality: 183716 bases at least Q30

Consensus quality: 187939 bases at least Q20

Insert size: 183000; agarose-fp

Insert size: 190857; sum-of-contigs

Quality coverage: 3.9 in Q20 bases; agarose-fp

Quality coverage: 3.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 29 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1 1628: contig of 1628 bp in length
* 1629 1728: gap of 100 bp
* 1729 2962: contig of 1234 bp in length
* 2963 3062: gap of 100 bp
* 3063 4983: contig of 1927 bp in length
* 4989 5089: gap of 100 bp
* 5090 6666: contig of 1577 bp in length
* 6667 8667: gap of 100 bp
* 8667 8667: contig of 1901 bp in length
* 8668 8767: gap of 100 bp
* 8768 11259: contig of 2492 bp in length
* 11260 11359: gap of 100 bp
* 11360 13557: contig of 2198 bp in length
* 13558 13657: gap of 100 bp
* 13658 17173: contig of 3516 bp in length
* 17174 17274: gap of 100 bp
* 17274 20060: contig of 2787 bp in length
* 20061 23843: gap of 100 bp
* 23844 23944: contig of 3683 bp in length
* 23944 27203: contig of 3260 bp in length
* 27204 27303: gap of 100 bp
* 27304 31592: contig of 4289 bp in length
* 31593 31692: gap of 100 bp
* 31693 35740: contig of 5048 bp in length
* 35741 36841: gap of 100 bp
* 36841 42197: contig of 3357 bp in length
* 42198 42297: gap of 100 bp
* 42298 47652: contig of 5355 bp in length
* 47653 47753: gap of 100 bp
* 47753 52307: contig of 4555 bp in length
* 52308 52407: gap of 100 bp
* 52408 58606: contig of 6199 bp in length
* 58607 65584: gap of 100 bp
* 65585 72492: contig of 6878 bp in length
* 72493 72592: gap of 100 bp
* 72593 78757: contig of 6164 bp in length
* 78757 85474: gap of 100 bp
* 85475 85574: contig of 6618 bp in length
* 85575 91951: gap of 100 bp
* 91952 92051: contig of 6377 bp in length
* 92052 99044: contig of 6993 bp in length
* 99045 99144: gap of 100 bp
* 99145 104884: contig of 5740 bp in length
* 104885 112130: gap of 100 bp
* 112130: contig of 7146 bp in length

* 112131 112230: gap of 100 bp
* 112231 118017: contig of 5787 bp in length
* 118018 126319: gap of 100 bp
* 126320 126419: gap of 100 bp
* 126420 137651: contig of 11232 bp in length
* 137652 137751: gap of 100 bp
* 137752 193697: contig of 55946 bp in length.

Location/Qualifiers

source

1..193697
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="RP11-200K12"
/clone_lib="RPC1-11 Human Male BAC"

misc_feature

1..1628
/note="assembly_fragment"
1729..2962
/note="assembly_fragment"
3063..4989
/note="assembly_fragment"
5090..5666
/note="assembly_fragment"
6767..8667
/note="assembly_fragment"
8768..11259
/note="assembly_fragment"
11360..13557
/note="assembly_fragment"
13658..17173
/note="assembly_fragment"
17274..20060
/note="assembly_fragment"
20161..23843
/note="assembly_fragment"
23944..27203
/note="assembly_fragment"
27304..31592
/note="assembly_fragment"
31693..36740
/note="assembly_fragment"
36841..42197
/note="assembly_fragment"
42298..47652
/note="assembly_fragment"
47753..52307
/note="assembly_fragment"
clone_end:SP6
vector_side:right
52408..58606
/note="assembly_fragment"
58707..65584
/note="assembly_fragment"
65685..72492
/note="assembly_fragment"
72593..78756
/note="assembly_fragment"
78857..85474
/note="assembly_fragment"
85575..91951
/note="assembly_fragment"
92052..99044
/note="assembly_fragment"
99145..104884
/note="assembly_fragment"
104985..112130
/note="assembly_fragment"
112231..118017
/note="assembly_fragment"
118118..126319
/note="assembly_fragment"
clone_end:T7
vector_side:right

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1615
Center clone name: 440 B.20

```

* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

*	1	691:	contig of 691 bp in length
*	692	791:	gap of 100 bp
*	792	1505:	contig of 714 bp in length
*	1506	1605:	gap of 100 bp
*	1606	2290:	contig of 685 bp in length
*	2291	2390:	gap of 100 bp
*	2391	3097:	contig of 707 bp in length
*	3098	3197:	gap of 100 bp
*	3198	3998:	contig of 701 bp in length
*	3999	3998:	gap of 100 bp
*	3999	4681:	contig of 683 bp in length
*	4682	4781:	gap of 100 bp
*	4782	5452:	contig of 671 bp in length
*	5453	5552:	gap of 100 bp
*	5553	6286:	contig of 734 bp in length
*	6287	6386:	gap of 100 bp
*	6387	7117:	contig of 731 bp in length
*	7118	7217:	gap of 100 bp
*	7218	7935:	contig of 718 bp in length
*	7936	8035:	gap of 100 bp
*	8036	8734:	contig of 699 bp in length
*	8735	8834:	gap of 100 bp
*	8835	9537:	contig of 703 bp in length
*	9538	9637:	gap of 100 bp
*	9638	10318:	contig of 681 bp in length
*	10319	10418:	gap of 100 bp
*	10419	11120:	contig of 702 bp in length
*	11121	11220:	gap of 100 bp
*	11221	11920:	contig of 680 bp in length
*	11901	12000:	gap of 100 bp
*	12001	12700:	contig of 700 bp in length
*	12701	12800:	gap of 100 bp
*	12801	13493:	contig of 693 bp in length
*	13494	13593:	gap of 100 bp
*	13594	14282:	contig of 689 bp in length
*	14283	14382:	gap of 100 bp
*	14383	15074:	contig of 692 bp in length
*	15075	15174:	gap of 100 bp
*	15175	15847:	contig of 673 bp in length
*	15848	15947:	gap of 100 bp
*	15948	16225:	contig of 678 bp in length
*	16226	16725:	gap of 100 bp
*	16726	17417:	contig of 692 bp in length
*	17418	17517:	gap of 100 bp
*	17518	18211:	contig of 694 bp in length
*	18212	18311:	gap of 100 bp
*	18312	18996:	contig of 685 bp in length
*	18997	19096:	gap of 100 bp
*	19097	19756:	contig of 660 bp in length
*	19757	19856:	gap of 100 bp
*	19857	20563:	contig of 707 bp in length
*	20564	20663:	gap of 100 bp
*	20664	21383:	contig of 720 bp in length
*	21384	21483:	gap of 100 bp

[illegible]

RESULT	8
AC090565	
LOCUS	61995 bp DNA linear HTG 03-MAR-2001
DEFINITION	Homo sapiens chromosome 8 clone RP11-440B20 map 8, LOW-PASS SEQUENCE SAMPLING.
ACCESSION	AC090565
VERSION	AC090565.1 GI:13194295
KEYWORDS	HTG; HTGS PHASEO.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
Eukaryota;	Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia;	Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 61995)	
Barren,B., Linton,L., Nusbaum,C. and Lander,E.	
Title	Homo sapiens chromosome 8, clone RP11-440B20
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 61995)
AUTHORS	Barren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,

Barron, E., Banton, V., Boguslavsky, L., Boukhalter, B., Brown, A.,
Barna, N., Bastien, J., Benoit, R., Berthiaume, M., Blais, D.,
Camarda, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArelano, K., Dewar, K., Diaz, J., S.,
Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,
Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Hearford, A., Horton, U., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,
Lehocsky, J., Levine, R., Liu, G., MacLean, C., MacDonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retra, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Sognuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Travers, M., Travis, N., Triggilo, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.-J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE
 JOURNAL
 Submitted (03-MAR-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Direct Submission
 Zembek, D., Zimmer, A. and Zody, M.

```
* 21484 22188: contig of 705 bp in length
* 22189 22288: gap of 100 bp
* 22289 22977: contig of 689 bp in length
* 22978 23077: gap of 100 bp
* 23078 23780: contig of 703 bp in length
* 23781 23880: gap of 100 bp
* 23881 24548: contig of 668 bp in length
* 24549 24648: gap of 100 bp
* 24649 25349: contig of 701 bp in length
* 25350 25449: gap of 100 bp
* 25450 26126: contig of 677 bp in length
* 26127 26226: gap of 100 bp
* 26227 26894: contig of 668 bp in length
* 26895 26995: gap of 100 bp
* 26996 27691: contig of 696 bp in length
* 27692 27791: gap of 100 bp
* 27792 28495: contig of 705 bp in length
* 28496 28595: gap of 100 bp
* 28596 29294: contig of 699 bp in length
* 29295 29395: gap of 100 bp
* 29396 30105: contig of 711 bp in length
* 30106 30205: gap of 100 bp
* 30206 30898: contig of 693 bp in length
* 30899 30999: gap of 100 bp
* 30999 31707: contig of 708 bp in length
* 31708 31806: gap of 100 bp
* 31807 32486: contig of 680 bp in length
* 32487 32586: gap of 100 bp
* 32587 33285: contig of 699 bp in length
* 33286 33385: gap of 100 bp
* 33386 34124: contig of 739 bp in length
* 34125 34225: gap of 100 bp
* 34226 34905: contig of 681 bp in length
* 34906 35005: gap of 100 bp
* 35006 35692: contig of 687 bp in length
* 35693 35792: gap of 100 bp
* 35793 36481: contig of 689 bp in length
* 36482 36581: gap of 100 bp
* 36582 37296: contig of 715 bp in length
* 37297 37396: gap of 100 bp
* 37397 38103: contig of 707 bp in length
* 38104 38203: gap of 100 bp
* 38204 38921: contig of 718 bp in length
* 38922 39021: gap of 100 bp
* 39022 39713: contig of 698 bp in length
* 39714 39820: gap of 100 bp
* 39821 40493: contig of 674 bp in length
* 40494 40593: gap of 100 bp
* 40594 41299: contig of 706 bp in length
* 41300 41399: gap of 100 bp
* 41400 42111: contig of 712 bp in length
* 42112 42211: gap of 100 bp
* 42212 42910: contig of 699 bp in length
* 42911 43010: gap of 100 bp
* 43011 43680: contig of 670 bp in length
* 43681 43780: gap of 100 bp
* 43781 44498: contig of 718 bp in length
* 44499 44599: gap of 100 bp
* 44599 45303: contig of 705 bp in length
* 45304 45404: gap of 100 bp
* 45404 46120: contig of 717 bp in length
* 46121 46220: gap of 100 bp
* 46221 46928: contig of 708 bp in length
* 46929 47029: gap of 100 bp
* 47030 47733: contig of 705 bp in length
* 47734 48526: contig of 693 bp in length
* 48527 48626: gap of 100 bp
* 48627 49335: contig of 709 bp in length
* 49336 49436: gap of 100 bp
* 49436 50108: contig of 673 bp in length
* 50109 50209: gap of 100 bp
* 50209 50894: contig of 686 bp in length

Query Match 8.6%; Score 201; DB 2; Length 61395;
Best Local Similarity 100.0%; Pred. No. 2.6e-108;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1408 GTTGCTGCTGTTGTCAGTAATCGGAAGGACAAATACGATGCTATTAAAAATACCTG 1467
Db 22689 GTTGCTGCTGTTGTCAGTAATCGGAAGGACAAATACGATGCTATTAAAAATACCTG 22748

QY 1468 TGTACAGATTGCCCTACCCCAAGTCAGTGTGGTGGCCGACCTTAGGCAACAGCAA 1527
Db 22749 TGTACAGATTGCCCTACCCCAAGTCAGTGTGGTGGCCGACCTTAGGCAACAGCAA 22808

QY 1528 ACTGTCATGCCCATTTGCTACAAAGATTGCCCTACAGATGAAGTCAAGATGGAGGAGAG 1587
Db 22809 ACTGTCATGCCCATTTGCTACAAAGATTGCCCTACAGATGAAGTCAAGATGGAGGAGAG 22868

QY 1588 CTCTGAGGGTGGACATCCCC 1608
Db 22869 CTCTGAGGGTGGACATCCCC 22889

RESULT 9
AC090147/c
LOCUS 66611 bp DNA linear HTG 18-MAY-2001
DEFINITION Homo sapiens chromosome 8 clone RP11-382A18 map 8, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC090147
VERSION AC090147.3 GI:14141769
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 66611)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 8, clone RP11-382A18
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 66611)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavskiy, L., Boukhgalter, B., Brown, A.,
Camrata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J.,
Gardyna, S., Girde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,
Lehotzky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheters, R., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
TITLE Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
```

COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 18, 2001 this sequence version replaced gi:13324816.
 All repeats were identified using RepeatMasker:

Smits, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L11702
 Center clone name: 382_A_18

* NOTE: This record contains 80 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely for
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1	738:	contig of 738 bp in length
*	739	gap of 100 bp
*	838:	contig of 745 bp in length
*	839	gap of 100 bp
*	1584	gap of 100 bp
*	1684	contig of 745 bp in length
*	2428:	gap of 100 bp
*	2528:	contig of 719 bp in length
*	3247:	contig of 100 bp
*	3248	gap of 100 bp
*	3348	contig of 732 bp in length
*	4080	gap of 100 bp
*	4180	contig of 709 bp in length
*	4888:	gap of 100 bp
*	4889	contig of 729 bp in length
*	4989	gap of 100 bp
*	5718	contig of 754 bp in length
*	5818	gap of 100 bp
*	6572	contig of 753 bp in length
*	6672	gap of 100 bp
*	7424:	contig of 736 bp in length
*	7425	gap of 100 bp
*	8260:	contig of 734 bp in length
*	8360:	gap of 100 bp
*	8361	contig of 734 bp in length
*	9094:	gap of 100 bp
*	9194:	contig of 720 bp in length
*	9195	gap of 100 bp
*	9915	contig of 725 bp in length
*	10015	gap of 100 bp
*	10739:	contig of 728 bp in length
*	10740	gap of 100 bp
*	10840	contig of 728 bp in length
*	11567:	gap of 100 bp
*	11568	contig of 723 bp in length
*	11668	gap of 100 bp
*	12391	contig of 728 bp in length
*	12491	gap of 100 bp
*	13218:	contig of 703 bp in length
*	13219	gap of 100 bp
*	13319	contig of 725 bp in length
*	14021:	contig of 737 bp in length
*	14121:	gap of 100 bp
*	14846:	contig of 737 bp in length
*	14847	gap of 100 bp
*	14947	contig of 743 bp in length
*	15684	gap of 100 bp
*	15784	contig of 743 bp in length
*	16526:	gap of 100 bp
*	16527	contig of 731 bp in length
*	16627	gap of 100 bp
*	17358	contig of 739 bp in length
*	17458	gap of 100 bp
*	18197	contig of 719 bp in length
*	18297	gap of 100 bp
*	19016	contig of 741 bp in length
*	19116	gap of 100 bp
*	19857	contig of 744 bp in length
*	19857	gap of 100 bp
*	20678:	contig of 722 bp in length
*	20778:	gap of 100 bp
*	21510:	contig of 732 bp in length
*	21610:	gap of 100 bp
*	22411:	contig of 731 bp in length
*	22441:	gap of 100 bp
*	23216:	contig of 775 bp in length
*	23316:	gap of 100 bp
*	24047:	contig of 731 bp in length
*	24147:	gap of 100 bp
*	24922:	contig of 775 bp in length
*	25022:	gap of 100 bp
*	25772:	contig of 750 bp in length
*	25872:	gap of 100 bp
*	26583:	contig of 711 bp in length
*	26683:	gap of 100 bp
*	27416:	contig of 733 bp in length
*	27516:	gap of 100 bp
*	28445:	contig of 729 bp in length
*	28345:	gap of 100 bp
*	28346	contig of 739 bp in length
*	29085	gap of 100 bp
*	29184:	contig of 734 bp in length
*	29185	gap of 100 bp
*	30018:	contig of 742 bp in length
*	30760:	contig of 742 bp in length
*	30761	gap of 100 bp
*	31602:	contig of 742 bp in length
*	31702:	gap of 100 bp
*	32436:	contig of 734 bp in length
*	32437	gap of 100 bp
*	32537	contig of 718 bp in length
*	33255	gap of 100 bp
*	33355	contig of 756 bp in length
*	34111	gap of 100 bp
*	34211	contig of 738 bp in length
*	34948:	contig of 774 bp in length
*	35048:	gap of 100 bp
*	35822:	contig of 774 bp in length
*	35922:	gap of 100 bp
*	36618:	contig of 696 bp in length
*	36718:	gap of 100 bp
*	37451:	contig of 733 bp in length
*	37551:	gap of 100 bp
*	38279:	contig of 728 bp in length
*	38379:	gap of 100 bp
*	39086:	contig of 707 bp in length
*	39186:	gap of 100 bp
*	39928:	contig of 742 bp in length
*	40028:	gap of 100 bp
*	40763:	contig of 735 bp in length
*	40863:	gap of 100 bp
*	41614:	contig of 751 bp in length
*	41714:	gap of 100 bp
*	42444:	contig of 730 bp in length
*	42544:	gap of 100 bp
*	43298:	contig of 754 bp in length
*	43398:	gap of 100 bp
*	44101:	contig of 703 bp in length
*	44201:	gap of 100 bp
*	44940:	contig of 739 bp in length
*	45040:	gap of 100 bp
*	45737:	contig of 697 bp in length
*	45837:	gap of 100 bp
*	46560:	contig of 723 bp in length
*	46660:	gap of 100 bp
*	47394:	contig of 734 bp in length
*	47494:	gap of 100 bp
*	48223:	contig of 729 bp in length
*	48323:	gap of 100 bp
*	49063:	contig of 740 bp in length
*	49163:	gap of 100 bp
*	49903:	contig of 740 bp in length
*	50003:	gap of 100 bp
*	50044	contig of 744 bp in length

```
* 50748 50847: gap of 100 bp
* 50848 51606: contig of 759 bp in length
* 51607 51706: gap of 100 bp
* 52462 52562: contig of 756 bp in length
* 52563 53294: contig of 100 bp
* 53295 53394: contig of 732 bp in length
* 54138 54238: gap of 100 bp
* 54239 54956: contig of 718 bp in length
* 54957 55056: gap of 100 bp
* 55057 55791: contig of 735 bp in length
* 55792 55891: gap of 100 bp
* 55892 56624: contig of 733 bp in length
* 56625 56724: gap of 100 bp
* 56725 57456: contig of 732 bp in length

Query Match
Best Local Similarity 100.0%; Pred. No. 6.9e-65; Length 66611;
Matches 129; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1938 CCTAGACTAACGGTAATTTGGTGAAGAAAAGAGTGAACACAGATTTTCTCAGTC 1997
Db 15977 CCTAGACTAACGGTAATTTGGTGAAGAAAAGAGTGAACACAGATTTTCTCAGTC 15918

QY 1998 TGGAGGAGACTTCAGATCCACTCTCTGGAACAGTATTGATGTAGAGTTACCGACC 2057
Db 15917 TGGAGGAGACTTCAGATCCACTCTCTGGAACAGTATTGATGTAGAGTTACCGACC 15858

QY 2058 AGAATGGTA 2066
Db 15857 AGAATGGTA 15849

RESULT 10
AC090147 66611 bp DNA linear HTG 18-MAY-2001
LOCUS Homo sapiens chromosome 8 clone RP11-382A18 map 8, LOW-PASS
DEFINITION SEQUENCE SAMPLING.
ACCESSION AC090147
VERSION AC090147.3 GI:14141769
KEYWORDS HTG; HTGS PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
1 (bases 1 to 66611)
Homo sapiens chromosome 8, clone RP11-382A18
Unpublished
2 (bases 1 to 66611)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B., Brown,A.,
Camata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Fairo,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
```

```
JOURNAL Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 18, 2001 this sequence version replaced gi:13324816.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIER
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11702
Center clone name: 382_A_18
* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 738: contig of 738 bp in length
* 739 938: gap of 100 bp
* 839 1583: contig of 745 bp in length
* 1584 1683: gap of 100 bp
* 1684 2428: contig of 745 bp in length
* 2429 2528: gap of 100 bp
* 2529 3247: contig of 719 bp in length
* 3248 3348: gap of 100 bp
* 3348 4079: contig of 732 bp in length
* 4080 4179: gap of 100 bp
* 4180 4888: contig of 709 bp in length
* 4889 4988: gap of 100 bp
* 4989 5717: contig of 729 bp in length
* 5718 5817: gap of 100 bp
* 5818 6571: contig of 754 bp in length
* 6572 6671: gap of 100 bp
* 6672 7424: contig of 753 bp in length
* 7425 7524: gap of 100 bp
* 7525 8260: contig of 736 bp in length
* 8261 8361: gap of 100 bp
* 8361 9095: contig of 734 bp in length
* 9095 9194: gap of 100 bp
* 9194 9914: contig of 720 bp in length
* 9915 10014: gap of 100 bp
* 10014 10739: contig of 725 bp in length
* 10740 10839: gap of 100 bp
* 10840 11567: contig of 728 bp in length
* 11568 11667: gap of 100 bp
* 11668 12390: contig of 723 bp in length
* 12391 12490: gap of 100 bp
* 12491 13218: contig of 728 bp in length
* 13219 13318: gap of 100 bp
* 13319 14021: contig of 703 bp in length
* 14022 14121: gap of 100 bp
* 14122 14846: contig of 725 bp in length
* 14847 14946: gap of 100 bp
* 14947 15683: contig of 737 bp in length
* 15684 15783: gap of 100 bp
* 15784 16526: contig of 743 bp in length
* 16527 16626: gap of 100 bp
* 16627 17357: contig of 731 bp in length
* 17358 17457: gap of 100 bp
* 17458 18136: contig of 739 bp in length
* 18137 18296: gap of 100 bp
* 18297 19015: contig of 719 bp in length
* 19016 19115: gap of 100 bp
* 19116 19856: contig of 741 bp in length
```

Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 18, 2001 this sequence version replaced gi:13324816.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIER
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11702
Center clone name: 382_A_18

* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```
* 19557 19556: gap of 100 bp
* 19557 20678: contig of 722 bp in length
* 20678 20778: gap of 100 bp
* 20778 21510: contig of 732 bp in length
* 21510 21610: gap of 100 bp
* 21610 22341: contig of 731 bp in length
* 22341 22441: gap of 100 bp
* 22441 23216: contig of 775 bp in length
* 23216 23317: gap of 100 bp
* 23317 24047: contig of 731 bp in length
* 24047 24147: gap of 100 bp
* 24147 24922: contig of 775 bp in length
* 24922 25022: gap of 100 bp
* 25022 25772: contig of 750 bp in length
* 25772 25872: gap of 100 bp
* 25872 26583: contig of 711 bp in length
* 26583 26683: gap of 100 bp
* 26683 27416: contig of 733 bp in length
* 27416 27516: gap of 100 bp
* 27516 28245: contig of 729 bp in length
* 28245 28345: gap of 100 bp
* 28345 29084: contig of 739 bp in length
* 29084 29184: gap of 100 bp
* 29184 29919: contig of 734 bp in length
* 29919 30760: contig of 742 bp in length
* 30760 30860: gap of 100 bp
* 30860 31702: contig of 742 bp in length
* 31702 32436: contig of 734 bp in length
* 32436 32536: gap of 100 bp
* 32536 33254: contig of 718 bp in length
* 33254 33354: gap of 100 bp
* 33354 34110: contig of 756 bp in length
* 34110 34210: gap of 100 bp
* 34210 34948: contig of 738 bp in length
* 34948 35048: gap of 100 bp
* 35048 35822: contig of 774 bp in length
* 35822 35922: gap of 100 bp
* 35922 36618: contig of 696 bp in length
* 36618 36719: gap of 100 bp
* 36719 37451: contig of 733 bp in length
* 37451 37551: gap of 100 bp
* 37551 38279: contig of 728 bp in length
* 38279 38379: gap of 100 bp
* 38379 39086: contig of 707 bp in length
* 39086 39186: gap of 100 bp
* 39186 39288: contig of 742 bp in length
* 39288 40028: gap of 100 bp
* 40028 40763: contig of 735 bp in length
* 40763 40863: gap of 100 bp
* 40863 41614: contig of 751 bp in length
* 41614 41714: gap of 100 bp
* 41714 42444: contig of 730 bp in length
* 42444 42544: gap of 100 bp
* 42544 43298: contig of 754 bp in length
* 43298 43398: gap of 100 bp
* 43398 44101: contig of 703 bp in length
* 44101 44201: gap of 100 bp
* 44201 44940: contig of 739 bp in length
* 44940 45040: gap of 100 bp
* 45040 45737: contig of 697 bp in length
* 45737 45837: gap of 100 bp
* 45837 46560: contig of 723 bp in length
* 46560 46660: gap of 100 bp
* 46660 47394: contig of 734 bp in length
* 47394 47494: gap of 100 bp
* 47494 48223: contig of 729 bp in length
* 48223 48323: gap of 100 bp
* 48323 49063: contig of 740 bp in length
* 49063 49163: gap of 100 bp
* 49163 49904: contig of 740 bp in length
* 50003: gap of 100 bp

* 50004 50747: contig of 744 bp in length
* 50747 50847: gap of 100 bp
* 50847 51606: contig of 759 bp in length
* 51606 51706: gap of 100 bp
* 51706 52462: contig of 756 bp in length
* 52462 52562: gap of 100 bp
* 52562 53294: contig of 732 bp in length
* 53294 53394: gap of 100 bp
* 53394 54138: contig of 744 bp in length
* 54138 54238: gap of 100 bp
* 54238 54956: contig of 718 bp in length
* 54956 55056: gap of 100 bp
* 55056 55791: contig of 735 bp in length
* 55791 55891: gap of 100 bp
* 55891 56624: contig of 733 bp in length
* 56624 56724: gap of 100 bp
* 56724 57456: contig of 732 bp in length
* 57456 57525: gap of 100 bp

Query Match 5.4%; Score 125; DB.2; Length 66611;
Best Local Similarity 100.0%; Fred. No. 1.8e-62;
Matches 125; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 AGGTTACTGAAGTTTGTAGTAAGACCCGGAATGGAGAGATGTGAGGATAAGCATCATT 331
Db 45398 AGGTTACTGAAGTTTGTAGTAAGACCCGGAATGGAGAGATGTGAGGATAAGCATCATT 45457

QY 332 TAACAAATGAATTCACCTACATACACCAACTGTTGTCAGTCTATATATATTTTCA 391
Db 45458 TAACAAATGAATTCACCTACATACACCAACTGTTGTCAGTCTATATATATTTTCA 45517

QY 392 GGAGG 396
Db 45518 GGAGG 45522

RESULT 11
BD114143
LOCUS BD114143 461 bp DNA linear PAT 18-SEP-2002
DEFINITION EST and encoded human protein.
ACCESSION BD114143
VERSION BD114143.1 GI:23209047
KEYWORDS JP 2002010789-A/6220.
SOURCE Homo sapiens (human)
. ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 461)
Edwards, J.B.D.M., Jobert, S. and Giordano, J.E.
EST and encoded human protein
Patent: JP 2002010789-A 6220 15-JAN-2002;
GENSET CORP
OS Homo sapiens (human)
PN JP 2002010789-A/6220
PD 15-JAN-2002
PF 07-AUG-2000 JP 2000280989
PR 05-AUG-1999 US 60/147499
PI JEAN EAPUTIST DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EVE PI
GIORDANO
PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC
C12N1/21,
PC C12N5/10, C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC
C12N15/00
CC n=a, g, c or t Location/Qualifiers
FH Key Location/Qualifiers
FT misc feature 107.
FEATURES
source
1..461
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 136 a 107 c 129 g 87 t 2 others
ORIGIN
```

FEATURES	source
Location/Qualifiers	
1..3910	
/organism="Mus musculus"	
/mol_type="mRNA"	
/db_xref="taxon:10090"	
/map="5p distal-G2 proximal"	
/tissue_lib="testis"	
1..3910	
/gene="miwi"	
156..2744	
/gene="miwi"	
/note="piwi family"	
/codon_start=1	
/product="MIWI"	
/protein_id="BAA93705.1"	
/db_xref="GI:7416111"	
/translation="MTGRARARGRARGOETVQHVGAASQOQYIPPRPOOSPTEG DLVGRGORGWVGATSKSQELOISAGFOELSLAERGRRRDPHDLGVNTRQMLDVK BSKTGSSGIIYKLSINHERLTSRPOWALYOXHIDYELMEARRLSALLFOHEDLIG CHAFDTFLPDKRLOHKVTEVPSOTRGEHVRITITLNELPPTSPCTCLOFYNIJFR RLTKIMNQOQGRNYINSPSDPIDINHRILVTPGFTTSLQIYENIMLCTIDVSHKVL SETVLDPMFYLYQTEEHKFOVSKELIGLIVLTKYNNKTYRVDIDWDQPKSTFK KADGSEVFLFYRYQYQNEILTDLQPVLSQVFKRRGPGGTLPGPAMLIPELCYLT LTKDMNRPNWMDLVAHRTLPEQORQEVGLIDYIHKDDNVQRELRDWMGLSFDNL LSPSGRIIQSEKIHOGGKTEDYNPOFADWSKEATGPLISVKPELDNWLIIYTRNVEA ANSLQNLKFTVPANGIQWKKALMEVDRTTEAYLRALQOKVTSPTQIVVCLSSNRK DKYLAIKYLCTDCTPSCVVARILGKQVNAIAKIALQONCKMGELWLRVMDPL KLAMTVGDCVHDTAGRSIAGFVASINEGTRMFSRCVFDQGOELVDLGVKVCLOA ALRAMSGNEVMPYSIVIVYDVGQGLKTLVNYEVPQFDCGLKSVGRGYNPRLTIVV VKRYNARFPAQSGRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSQSVSPHYNVI YDSSGLKPDHIOQLIYKLVCHVYNNPVGIVRVPAPQCYAHKLAFLVQGSIHREPNSLS NKLIIYI"	
BASE COUNT	993 a 970 c 1063 g 884 t
ORIGIN	
Query Match	1.5%; Score 35; DB 10; Length 3910;
Best Local Similarity	100.0%; Pred. No. 3.1e-08;
Matches	35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1687 GCCAGCATCAATGAAGGATGACCCGCTGTTCTC 1721
DB	2103 GCCAGCATCAATGAAGGATGACCCGCTGTTCTC 2137
RESULT 14	
AF438405	
LOCUS	Mus musculus MIWI mRNA, complete cds.
DEFINITION	4064 bp mRNA linear ROD 13-NOV-2001
ACCESSION	AF438405
VERSION	AF438405.1 GI:16905060
KEYWORDS	
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	Deng, W. and Lin, H.
JOURNAL	Miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 4064)
TITLE	Deng, W. and Lin, H.
JOURNAL	Submitted (22-OCT-2001) Department of Cell Biology, Duke University, Room 412, Nanaline Duke Building, Durham, NC 27710, USA
FEATURES	
source	Location/Qualifiers
	1..4064
	/organism="Mus musculus"
	/mol_type="mRNA"
	/db_xref="taxon:10090"
	/chromosome="5"
	/tissue_type="testis"
	191..2779
CDS	

```
/note="Piwi-like protein"
/codon_start=1
/product="MIWI"
/protein_id="AAL31014.1"
/db_xref="GI:16905061"
/translation="MTGRARARARGAQETVHVGAASQPGYIPRPOOSPTEG
DLVGRGRQMVVGTAKSQELQISAGFELSIAERGGRRDRPHDLGVNTRQNLDHVX
ESKTSSGIIVLKLSTHFLRTPQWALYQYHIDYNPLMEARLRALLFQHEDLIGR
CHAFDGTIILFLKRLQKVTVEFSRNGEHRVITITLNELPSTPTCLOFYNIIFR
RLIKMNLQOIGNYNPSDIDIPNHLVINPGFTTIIQYENIMLCTDVSHKVLIR
SEVLDFMFLYQOYEEHFKQEOVSKEILGLI VLTKNKTYRVDIDWDONPKSTFK
KADGSEVSELEYRYKQYNOBEITLQKPLVSPKRRRGGTLPGEAMLI BELCVLTG
LTDKMRDNFVMDLAVHITLLEEQREVRGLDITIHKDDNVQRELRWGLSFDNHL
LSFSRLIQLSEKIHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNTEA
ANSILQNLFEVTPANGIQMKKAIMIEYDRTETAYLRALQOKVTSDTQIVVCLSSNRK
KYDAIKYLCVCTPSCQVARTICQOCTVMAIATKIALQNMCKMGELMRVDMPL
KLAMIVIGDYCHYTTAGRRSIAGFVASINEGTRFNSRVCVFDGQELVDGLKVCLOA
ALAWSGCNEYMPSRVIYVDGQGLKTLVNYEVPQELDLCKLSVGRGVNPELTIV
VKRVARFPFAQGGRLQNPGLPGTVIDVEVTRPEWIDFFIVSOAVRSQSVSPHYNVI
YDSGLKPDHIOIRLTLYKLCHVYNNPVGIVRVPAPQYAHKLAFLVQSIHREPNLSLS
NRLYYL"
```

```
BASE COUNT 1115 a 982 c 1077 g 890 t
ORIGIN
Query Match 1.5%; Score 35; DB 10; Length 4064;
Best Local Similarity 100.0%; Pred. No. 3.1e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1687 GCCAGCATCAATGAAGGATGACCCGCTGGTTCTC 1721
Db 2138 GCCAGCATCAATGAAGGATGACCCGCTGGTTCTC 2172

RESULT 15
AX39581.9 501 bp DNA linear PAT 18-MAY-2002
LOCUS AX395819
DEFINITION Sequence 34 from Patent WO0212328.
ACCESSION AX395819
VERSION AX395819.1 GI:21066566
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
1
REFERENCE
AUTHORS King, G.E., Meagher, M.J., Xu, J. and Secrist, H.
TITLE Compositions and methods for the therapy and diagnosis of colon
cancer
JOURNAL Patent: WO 0212328-A 34 14-FEB-2002;
CORIXA CORPORATION (US)
FEATURES
source
1. .501
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 134 a 81 c 93 g 193 t
ORIGIN
```

```
Query Match 1.5%; Score 34; DB 6; Length 501;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2271 CCAGAGTATTACAGAGAGCCAAATCTCTCACTG 2304
Db 1 CCAGAGTATTACAGAGAGCCAAATCTCTCACTG 34
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2003, 15:49:54 ; Search time 616 Seconds

(without alignments)
10201.772 Million cell updates/sec

Title: US-10-043-774B-1

Perfect score: 2328

Sequence: 1 atgatcttgggtgaacac.....accgcctttactactcttaa 2328

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N Geneseq 19Jun03:*

```

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
25: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2042	87.7	3472	21 AAA07588	Human piwi gene, d
2	1569	67.4	2292	24 ABV89287	Human colon cancer
3	316	13.6	367	25 ABZ19511	Group III cDNA can
4	284	12.2	300	20 AAZ13404	Human gene express
5	269	11.6	423	21 AAF21880	Human breast and o
6	214	9.2	367	25 ABZ20005	Group III cDNA can
7	158	6.8	209	25 ABZ19102	Group III cDNA can
8	60	2.6	60	24 ABN42800	Human spliced tran

c	9	58	2.5	460	22	AA525525	Human ovarian PCR-
	10	35	1.5	4064	21	AAA07587	Mouse piwi gene, d
	11	34	1.5	501	24	ABK44483	cDNA encoding colo
c	12	22	0.9	22	25	ABX15293	ASM 698-13 B.1 PCR
	13	21	0.9	1664976	19	AAV21209	Methanococcus jann
c	14	19	0.8	302	24	ABN23657	ASM 698-13 B.1 PCR
c	15	19	0.8	1061	21	AAAC98189	Human ORFX polynuc
c	16	19	0.8	1342	21	AAAC78014	Human colon cancer
c	17	19	0.8	1732	13	AAQ29914	Human cancer assoc
	18	19	0.8	1821	24	ABZ12729	Helianthin gene
c	19	19	0.8	1850	23	AA567089	Arabidopsis thalia
c	20	19	0.8	2288	22	AAI61070	DNA encoding novel
c	21	19	0.8	2362	21	AAAC77324	Human polynucleoti
c	22	19	0.8	2371	22	AAI59284	Human ORFX ORF2879
c	23	19	0.8	2412	25	ACA04715	Human polynucleoti
c	24	19	0.8	2432	22	AAH15530	cDNA encoding huma
c	25	19	0.8	2636	22	AAF55426	Human cDNA sequenc
c	26	19	0.8	3524	23	ABL07298	Nucleotide sequenc
c	27	19	0.8	5556	24	AAI47132	Drosophila melanog
c	28	19	0.8	6461	22	AA500033	Pyrin domain conta
c	29	19	0.8	6693	23	ABL07428	Human ATLAS-2-enco
c	30	19	0.8	8093	22	AAK86634	Drosophila melanog
c	31	19	0.8	218336	25	AAQ76678	Human immune/haema
c	32	19	0.8	349980	22	AAH41224	Androgen receptor
c	33	19	0.8	36	21	AAA35613	Pyrococcus abyssi
	34	18	0.8	36	21	AAA35613	Permuten linker e
	35	18	0.8	36	21	AAA35813	Permuten linker e
c	36	18	0.8	36	21	AAA38114	Permuten linker e
c	37	18	0.8	36	21	AAA38114	Human granulocyte
c	38	18	0.8	50	22	ABZ00670	Human leukocyte ge
	39	18	0.8	57	22	AAZ29709	Human Ig leader se
c	40	18	0.8	65	24	ABN52306	Mouse spliced tran
c	41	18	0.8	140	21	AAAC31484	Human secreted pro
c	42	18	0.8	244	22	AA540080	DNA encoding human
	43	18	0.8	278	23	ABV61749	Human prostate exp
	44	18	0.8	306	24	ABN26826	Human ORFX polynuc
	45	18	0.8				

ALIGNMENTS

```

RESULT 1
ID AAA07588 standard; DNA; 3472 BP.
XX
XX AAA07588;
XX
XX 29-AUG-2000 (first entry)
XX
XX Human piwi gene, designated hiwi.
XX
XX Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
XX anaemia; immunodeficiency; male infertility; human; ds.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 167..2752
XX FT /*tag= a
XX FT /product= hiwi
XX FT /transl_except= (pos:392..394; aa:Xaa)
XX FT /transl_except= (pos:1073..1075; aa:Xaa)
XX FT /transl_except= (pos:2369..2370; aa:Xaa)
XX FT /note= "Xaa= Leu or Ile"
XX
XX WO200032039-A1.
XX
XX 08-JUN-2000.
XX
XX 03-DEC-1999; 99WO-US28764.
XX
XX 04-DEC-1998; 98US-0110901.
XX

```

XX (UYDU-) UNIV DUKE.
XX PA
XX PI
XX Lin H;
XX WPI; 2000-412085/35.
DR P-PSDB; AAY90235.
XX
PT Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT therapy of diseases such as cancer and in various research and
XX diagnostic applications -
XX
PS Claim 19; Page 189-194; 201pp; English.
XX
CC This sequence encodes the human piwi family protein, designated
CC hiwi. The piwi family nucleic acids and polypeptides are used in gene
CC therapy of diseases such as cancer and also in various research and
CC diagnostic applications. The sequences can also be used to treat
CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
SQ Sequence 3472 BP; 1044 A; 694 C; 796 G; 935 T; 3 other;

Query Match 87.7%; Score 2042; DB 21; Length 3472;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2292; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 32 TAGACCATGTTAAAGAAATCAAAAACAGGTTCTTTCAGGCATTATAGTAAGGTTAAGCACTA 91
DB 456 TAGACCATGTTAAAGAAATCAAAAACAGGTTCTTTCAGGCATTATAGTAAGGTTAAGCACTA 515

QY 92 ACCATTCCGGCTGACATCCCGTCCCGAGTGGGCTTATATCATGATACATTTGACTATA 151
DB 516 ACCATTCCGGCTGACATCCCGTCCCGAGTGGGCTTATATCATGATACATTTGACTATA 575

QY 152 ACCACTGATGGAGCCAGAGACTCGTTCACTCTCTTTTCAACAGAGAGTCTAA 211
DB 576 ACCACTGATGGAGCCAGAGACTCGTTCACTCTCTTTTCAACAGAGAGTCTAA 635

QY 212 TTGAAAGTGTCTATGCTTTTGTGGAACGATATATTTTACCTTAAAGACTACAGCAA 271
DB 636 TTGAAAGTGTCTATGCTTTTGTGGAACGATATATTTTACCTTAAAGACTACAGCAA 695

QY 272 AGTTACTGAAGTTTTPAGTAAGACCCGGAATGGAGAGGATGTGAGATTAACGATCACTT 331
DB 696 AGTTACTGAAGTTTTPAGTAAGACCCGGAATGGAGAGGATGTGAGATTAACGATCACTT 755

QY 332 TAACAAATGAATTCACCTACATCAACCACTTTTGTGAGTCTATAATATTTTCA 391
DB 756 TAACAAATGAATTCACCTACATCAACCACTTTTGTGAGTCTATAATATTTTCA 815

QY 392 GGAGGCTTTTGAATATCATGAATTTGCAACAAATTTGGACGAAATTTATTAACCCAAATG 451
DB 816 GGAGGCTTTTGAATATCATGAATTTGCAACAAATTTGGACGAAATTTATTAACCCAAATG 875

QY 452 ACCCAATGATATTCAGATCAAGTTTGGTGTATTTGGCTTGGCTTCACTACTTCCATCC 511
DB 876 ACCCAATGATATTCAGATCAAGTTTGGTGTATTTGGCTTGGCTTCACTACTTCCATCC 935

QY 512 TTCAGTATGAACACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCTCTCGAAGTG 571
DB 936 TTCAGTATGAACACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCTCTCGAAGTG 995

QY 572 AGACTGTTTGGATTTTCACTTTTATCATCATGACGTTAGCCATAAAGTCTCTCGAAGTG 631
DB 996 AGACTGTTTGGATTTTCACTTTTATCATCATGACGTTAGCCATAAAGTCTCTCGAAGTG 1055

QY 632 AACAGTTTCCAAAGAACTAATAGTTTGTGTTTCTTTTCAAGTATTAACATTAAGCAT 691
DB 1056 AACAGTTTCCAAAGAACTAATAGTTTGTGTTTCTTTTACCAAGTATTAACATTAAGCAT 1115

QY 692 ACAGATGGATGATTTGACTGGGACCCAGATCCAGAGCAGCTTTTAAAGAACCCGAG 751
DB 1116 ACAGATGGATGATTTGACTGGGACCCAGATCCAGAGCAGCTTTTAAAGAACCCGAG 1175

QY 752 GCTCTGAAGTCAAGTCTTCTTGAATTAAGTCTAAGTCTAAGTCTAAGTCTAAGTCTAAGTCT 811
DB 1176 GCTCTGGGGTCAAGTCTTCTTAGAATACTAAGTCTAAGTCTAAGTCTAAGTCTAAGTCT 1235

QY 812 TGAGCAGCCTGCTCTTGGTCAAGCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 871
DB 1236 TGAGCAGCCTGCTCTTGGTCAAGCAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1295

QY 872 CAGGCGCTGCCATGCTCATTTCTGAGCTCTGCTATCTTACAGGCTCAATGATAAATG 931
DB 1296 CAGGCGCTGCCATGCTCATTTCTGAGCTCTGCTATCTTACAGGCTCAATGATAAATG 1355

QY 932 GTAATGATTTTAACTGATGAAGACTTAGCGCTTATCAAGACTAAGTCTAAGTCTAAGTCT 991
DB 1356 GTAATGATTTTAACTGATGAAGACTTAGCGCTTATCAAGACTAAGTCTAAGTCTAAGTCT 1415

QY 992 GGCAGCGTGAAGTGGAGAGCTCATTTGATTAATCATTAATAAAGAGAGAGAGAGAGAGAG 1051
DB 1416 GGCAGCGTGAAGTGGAGAGCTCATTTGATTAATCATTAATAAAGAGAGAGAGAGAGAG 1475

QY 1052 AGCTTCGAGACTGGGGTGTGAGCTTTGATTTCAACTTACTGCTCTTCTCAGGAAGATTT 1111
DB 1476 AGCTTCGAGACTGGGGTGTGAGCTTTGATTTCAACTTACTGCTCTTCTCAGGAAGATTT 1535

QY 1112 TCACAAAGAAAAGATTCACCAAGGTGGAAGAAACATTTGATTAATCAATCCACATTTGAG 1171
DB 1536 TCACAAAGAAAAGATTCACCAAGGTGGAAGAAACATTTGATTAATCAATCCACATTTGAG 1595

QY 1172 ATTGTCCTCAAGAAACAGAGAGTGCACCAATTAATTAAGTGTAAAGCCACTAGATTAAGT 1231
DB 1596 ATTGTCCTCAAGAAACAGAGAGTGCACCAATTAATTAAGTGTAAAGCCACTAGATTAAGT 1655

QY 1232 TGTGATCTATACGGGAGAAATTTAAGCAGCAAAATTCATTGATGATCAAAATCTATTTA 1291
DB 1656 TGTGATCTATACGGGAGAAATTTAAGCAGCAAAATTCATTGATGATCAAAATCTATTTA 1715

QY 1292 AAGTTACACAGCCATGGGCATGCAAAATGAGAAAGCAATTAATGATTAAGTGGATGACA 1351
DB 1716 AAGTTACACAGCCATGGGCATGCAAAATGAGAAAGCAATTAATGATTAAGTGGATGACA 1775

QY 1352 GAATGAGCCCTACTTAAAGTCTTACAGCAAAAGTCTACAGCAGACACCCAGATAGTTG 1411
DB 1776 GAATGAGCCCTACTTAAAGTCTTACAGCAAAAGTCTACAGCAGACACCCAGATAGTTG 1835

QY 1412 TCTGCTGTGTCAAGTAAATCGGAAGGACAAATACGATGCTATTAATAAATACTGTGTA 1471
DB 1836 TCTGCTGTGTCAAGTAAATCGGAAGGACAAATACGATGCTATTAATAAATACTGTGTA 1895

QY 1472 CAGATTGCCCTACCCCAAGTCTGAGTGTGGTGGCCCAAGCTTAGGCAACAGCAACTG 1531
DB 1896 CAGATTGCCCTACCCCAAGTCTGAGTGTGGTGGCCCAAGCTTAGGCAACAGCAACTG 1955

QY 1532 TCATGGCCATTTGCTACAAAGATTTGCCCTACAGATGAATGCAAGATGGGAGAGAGCTCT 1591
DB 1956 TCATGGCCATTTGCTACAAAGATTTGCCCTACAGATGAATGCAAGATGGGAGAGAGCTCT 2015

QY 1592 GGAAGGTGGACATCCCGCTGAAAGTCTGATGATGATGATGATGATGATGATGATGATGATG 1651
DB 2016 GGAAGGTGGACATCCCGCTGAAAGTCTGATGATGATGATGATGATGATGATGATGATG 2075

QY 1652 TGAACGCTGGGAGAGTCAATCGAGGATTTGTTGCCAGCATCAATGAGGATGAGCC 1711
DB 2076 TGAACGCTGGGAGAGTCAATCGAGGATTTGTTGCCAGCATCAATGAGGATGAGCC 2135

QY 1712 GCTGTTCTCAGCTGCATATTTTCAAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1771
DB 2136 GCTGTTCTCAGCTGCATATTTTCAAGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2195

QY 1772 TCTGCTCTGCAAGCGGCTCTGAGGCTTTGGAATAGCTGCAATGATGATGATGATGATGATG 1831
DB 2196 TCTGCTCTGCAAGCGGCTCTGAGGCTTTGGAATAGCTGCAATGATGATGATGATGATGATG 2255

QY 1832 TCATCGTGTACCGGATGCGGTAGGAGCGCGCAGCTGAAACACACTGGTGAACCTACGAG 1891
DB |||||
DB 2256 TCATCGTGTACCGGATGCGGTAGGAGCGCGCAGCTGAAACACACTGGTGAACCTACGAG 2315
QY 1892 TGCCACAGTTTTTGGATTGCTTAAATCCATTTGGTAGAGGTTACAACTAGACTAACGG 1951
DB |||||
DB 2316 TGCCACAGTTTTTGGATTGCTTAAATCCATTTGGTAGAGGTTACAACTAGACTAACGG 2375
QY 1952 TAATTGTGTGAAGAAAGAGTGAACACACAGATTTTTTGTCTGATGCTGGAGGAGCTTC 2011
DB |||||
DB 2376 TAATTGTGTGAAGAAAGAGTGAACACACAGATTTTTTGTCTGATGCTGGAGGAGCTTC 2435
QY 2012 AGAATCCATCTCTGAAACAGTTATTGATGATGATGATGATGATGATGATGATGATGAT 2071
DB |||||
DB 2436 AGAATCCATCTCTGAAACAGTTATTGATGATGATGATGATGATGATGATGATGATGAT 2495
QY 2072 TTTTATCTGTGAGCCAGGCTGTGAGAGTGGTAGTGTCTCCACACATTAACAATGTCA 2131
DB |||||
DB 2496 TTTTATCTGTGAGCCAGGCTGTGAGAGTGGTAGTGTCTCCACACATTAACAATGTCA 2555
QY 2132 TCTATGACAAACGCGGCTGAGCCAGCCAGCACATACAGCGCTTGACCTACAGCTGTGCC 2191
DB |||||
DB 2556 TCTATGACAAACGCGGCTGAGCCAGCCAGCACATACAGCGCTTGACCTACAGCTGTGCC 2615
QY 2192 ACATCTATTACAACTGCGCAGGCTGATTCGCTGCTCTGCTGCTGCTGCTGCTGCTGCTG 2251
DB |||||
DB 2616 ACATCTATTACAACTGCGCAGGCTGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2675
QY 2252 AGCTGGCTTTCTTGTGGCCAGAGTATTCACAGAGAGCCAAATCTGCTCTGCTGCTGCTG 2311
DB |||||
DB 2676 AGCTGGCTTTCTTGTGGCCAGAGTATTCACAGAGAGCCAAATCTGCTCTGCTGCTGCTG 2735
QY 2312 GCCTTTACTACCTCTAA 2328
DB |||||
DB 2736 GCCTTTACTACCTCTAA 2752

RESULT 2

ID ABV89287 standard; cDNA; 2292 BP.
AC ABV89287;
XX
XX 13-DEC-2002 (first entry).
DE Human colon cancer related cDNA SEQ ID NO 2602.
XX Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
XX SS.
OS Homo sapiens.
PN WO200258534-A2.
XX
PD 01-AUG-2002.
XX
XX 19-NOV-2001; 2001WO-US43704.
XX
XX 20-NOV-2000; 2000US-252222P.
XX 06-FEB-2001; 2001US-267011P.
XX 28-MAR-2001; 2001US-279670P.
XX 10-JUL-2001; 2001US-304037P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;
XX WPI; 2002-608400/65.
DR P-PSDB; ABP67995.
XX
XX New isolated tumor colon polynucleotide and polypeptide, useful for the
PT diagnosis, prevention and/or treatment of cancer, in particular colon
PT cancer -

XX Claim 1; SEQ ID NO 2602; 266pp + Sequence Listing; English.
PS
XX The invention relates to a human colon tumour expressed polynucleotide
CC (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
CC sequences that hybridize to (i), under moderately stringent conditions;
CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
CC degenerate variants of (i). The compositions and methods of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC cancer, particularly colon cancer. (i) can be used in gene therapy and
CC (ii) and (iii) are useful in pharmaceutical compositions such as vaccines.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 2292 BP; 692 A; 435 C; 512 G; 653 T; 0 other;
SQ
Query Match 67.4%; Score 1569; DB 24; Length 2292;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1569; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 760 GTCACTTCTTGAATATCTACAGGAGCAATCAACCAAGAGATCAACCACTTGAAGCAG 819
DB 4 GTCACTTCTTGAATATCTACAGGAGCAATCAACCAAGAGATCAACCACTTGAAGCAG 63
QY 820 CCTGTCTTGTTCAGCCAGCCAGAGAGGCGGCGCTGGGGGACACTGCCAGGCGCT 879
DB 64 CTTGTCTTGTTCAGCCAGCCAGAGAGGCGGCGCTGGGGGACACTGCCAGGCGCT 123
QY 880 GCCATGCTCATCTCTGAGCTCTGCTATCTTACAGGCTCTAACTGATAAAATGCGTAATGAT 939
DB 124 GCCATGCTCATCTCTGAGCTCTGCTATCTTACAGGCTCTAACTGATAAAATGCGTAATGAT 183
QY 940 TTTAAGCTGATGAAGACCTTAGCCGTTCTACAGACTAACTCCAGAGCAAGGCGAGCGT 999
DB 184 TTTAAGCTGATGAAGACCTTAGCCGTTCTACAGACTAACTCCAGAGCAAGGCGAGCGT 243
QY 1000 GAAGTGGGAGCACTCATTTGATTTACATTTCAATAAAGCAATATGTTCAAGGGAGCTTGA 1059
DB 244 GAAGTGGGAGCACTCATTTGATTTACATTTCAATAAAGCAATATGTTCAAGGGAGCTTGA 303
QY 1060 GACTGGGTTTGGAGCTTTGATTTCAACTTCTCTCTCAGAGAGAAATTTTGCAGAA 1119
DB 304 GACTGGGTTTGGAGCTTTGATTTCAACTTCTCTCTCAGAGAGAAATTTTGCAGAA 363
QY 1120 GAAAGATTCAACAGGTGGAAAAACAATTTGATTAACAATCCAAATTTGAGATTGTTCTC 1179
DB 364 GAAAGATTCAACAGGTGGAAAAACAATTTGATTAACAATCCAAATTTGAGATTGTTCTC 423
QY 1180 AAGAAACAAGAGGTGCAACCAATTAATGTTAAGCCCTAGATTAACCTGCTGTTGATC 1239
DB 424 AAGAAACAAGAGGTGCAACCAATTAATGTTAAGCCCTAGATTAACCTGCTGTTGATC 483
QY 1240 TATACGGAAGAAATTTATGAGCAGCAATTTCAATGATACAAAATCTATTAAAGTTACA 1299
DB 484 TATACGGAAGAAATTTATGAGCAGCAATTTCAATGATACAAAATCTATTAAAGTTACA 543
QY 1300 CCAGCCATGGGCAATGCAAAATGAGAAAGCAATTAATGATTAAGTGAATGACAGAACTGAA 1359
DB 544 CCAGCCATGGGCAATGCAAAATGAGAAAGCAATTAATGATTAAGTGAATGACAGAACTGAA 603
QY 1360 GCCTACTTAAGAGTCTTACGAGAAAGGTACAGCAGACACCCAGATAGTTGCTGCTG 1419
DB 604 GCCTACTTAAGAGTCTTACAGCAGAAAGGTACAGCAGACACCCAGATAGTTGCTGCTG 663
QY 1420 TTCTCAAGTAACTCGGAGGCAATTAATGATTAAGTGAATGACAGAACTGCTGATGCTG 1479
DB 664 TTCTCAAGTAACTCGGAGGCAATTAATGATTAAGTGAATGACAGAACTGCTGATGCTG 723
QY 1480 CCTACCCCAAGTCAAGTGTGTGTGGCGCCGCACTTAGGCCAAAACAGCAAACTGCTATGCC 1539
DB |||||

Db 724 CCTATCCCAAGTCAGTGTGTGTGTGGCCGGAACCTTAGGCAAAACAGCAAACTGTCTGGCC 783
Qy 1540 ATTGCTACAAAGATTGCCCTACAGATGAACCTGCAAGATGGGAGAGAGCTCTGGAGGGTG 1599
Db 784 ATTGCTACAAAGATTGCCCTACAGATGAACCTGCAAGATGGGAGAGAGCTCTGGAGGGTG 843
Qy 1600 GACATCCCTCTGAAGCTGTGATGATGCTGTGGCATCGATTGTACCATGATGACAGCT 1659
Db 844 GACATCCCTCTGAAGCTGTGATGATGCTGTGGCATCGATTGTACCATGATGACAGCT 903
Qy 1660 GGGGGAGGTCAATCGCAGATTGTTGCTCCAGATCAATGAAGGGATGACCCCTGGTTC 1719
Db 904 GGGCGGAGGTCAATCGCAGATTGTTGCTCCAGATCAATGAAGGGATGACCCCTGGTTC 963
Qy 1720 TCACGCTGCATATTTTCAGGATAGAGGACAGAGCTGTGATGGCTCAAAAGTCTGCCTG 1779
Db 964 TCACGCTGCATATTTTCAGGATAGAGGACAGAGCTGTGATGGCTCAAAAGTCTGCCTG 1023
Qy 1780 CAAGCGCTCTGAGGGCTTGGAAATAGCTGCAATGAGTACATGCCCGGATCATCGTG 1839
Db 1024 CAAGCGCTCTGAGGGCTTGGAAATAGCTGCAATGAGTACATGCCCGGATCATCGTG 1083
Qy 1840 TACCGGATCGGTAGAGAGCGCCAGCTGAAACACTGTGTGACTGACGAGTGCACAG 1899
Db 1084 TACCGGATCGGTAGAGAGCGCCAGCTGAAACACTGTGTGACTGACGAGTGCACAG 1143
Qy 1900 TTTTGGATTGCTTAAATCCATTGTAGAGGTACCAACCTAGACTAAACGGTAATTGTG 1959
Db 1144 TTTTGGATTGCTTAAATCCATTGTAGAGGTACCAACCTAGACTAAACGGTAATTGTG 1203
Qy 1960 GTGAAGAAAAGAGTGAACACAGATTTTTTGTCTCAGTCTGGAGGAAGCTTCAGATCCA 2019
Db 1204 GTGAAGAAAAGAGTGAACACAGATTTTTTGTCTCAGTCTGGAGGAAGCTTCAGATCCA 1263
Qy 2020 TTCTCTGGAACAGTATTGATGTAGAGGTACAGACAGAGTGTGATGACTTTTATC 2079
Db 1264 TTCTCTGGAACAGTATTGATGTAGAGGTACAGACAGAGTGTGATGACTTTTATC 1323
Qy 2080 GTGAGCCAGGCTGTGAGAGGTGTGTTTCTCCACACATTACAATGTCATCTATGAC 2139
Db 1324 GTGAGCCAGGCTGTGAGAGGTGTGTTTCTCCACACATTACAATGTCATCTATGAC 1383
Qy 2140 AACAGCGCTCTGAAGCCAGACCAATACAGCGGTGACCTACAGCTGTGACCACTAT 2199
Db 1384 AACAGCGCTCTGAAGCCAGACCAATACAGCGGTGACCTACAGCTGTGACCACTAT 1443
Qy 2200 TACAACTGGCCAGGTGTCTTCTGTTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 2259
Db 1444 TACAACTGGCCAGGTGTCTTCTGTTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1503
Qy 2260 TTTCTTGTGGCCAGAGTATTACAGAGAGGCCAAATCTGTCACTGTCAAAACCGCTTTAC 2319
Db 1504 TTTCTTGTGGCCAGAGTATTACAGAGAGGCCAAATCTGTCACTGTCAAAACCGCTTTAC 1563
Qy 2320 TACCTCTAA 2328
Db 1564 TACCTCTAA 1572

RESULT 3

ABZ19511
ID ABZ19511 standard; cDNA; 367 BP.

XX AC ABZ19511;

XX XX 23-JAN-2003 (first entry)

DE Group III cDNA cancer related clone SEQ ID NO:1937.

KW Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
KW immune response; virology; immunology; microbiology; molecular biology;
KW recombinant DNA technology; gene; ss.

OS Homo sapiens.
XX WO200278516-A2.
XX PD 10-OCT-2002.
XX PF 28-MAR-2002; 2002WO-US10421.
XX PR 30-MAR-2001; 2001US-280255P.
PR 28-AUG-2001; 2001US-315563P.
PR 09-JAN-2002; 2002US-347313P.
XX (CORI-) CORIXA CORP.
XX Wang T, Wang S, Bangur CS, Gaiger A;
XX WPI; 2003-058387/05.

XX New immunogenic polynucleotides or polypeptides useful for diagnosing,
PT preventing and treating cancer expressing CT or CP mRNA antigens, and
PT in virology, immunology, microbiology, molecular biology and
PT recombinant DNA techniques

PS Claim 1; SEQ ID 1937; 207pp; English.

XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
CC ABP5446 to ABP54472 represent protein (II) sequences, from the present
CC invention. (I) and (II) have cytostatic activity and can be used in gene
CC therapy and vaccines. (I), (II), antibodies and compositions from the
CC present invention are useful for diagnosing, preventing and treating
CC cancer, which expresses CT or CP mRNA antigens. They are useful for
CC stimulating immune response. They can also be useful in virology,
CC immunology, microbiology, molecular biology and recombinant DNA
CC techniques.

CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 367 BP; 118 A; 77 C; 91 G; 81 T; 0 other;

Query Match 13.6%; Score 316; DB 25; Length 367;
Best Local Similarity 99.7%; Pred. No. 1.9e-146;
Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1291 AAAGTTACACACCCATGGGCATGCAAAATGAGAAAAGCAATATGATTGAATGGATGAC 1350
Db 1 AAAGTTACACACCCATGGGCATGCAAAATGAGAAAAGCAATATGATTGAATGGATGAC 60

Qy 1351 AGAAGTGAAGCTTAAAGTCTTACAGCAAAAGTCTACAGACACCCAGATAGTT 1410
Db 61 AGAAGTGAAGCTTAAAGTCTTACAGCAAAAGTCTACAGACACCCAGATAGTT 120

Qy 1411 GTCTGTCTGTGTCAAGTAAATCGAAAGGCAAAATACGATCTATTAAAAATACCTGTGT 1470
Db 121 GTCTGTCTGTGTCAAGTAAATCGAAAGGCAAAATACGATCTATTAAAAATACCTGTGT 180

Qy 1471 ACAGATTGCCCTACCCCAAGTCTAGTGTGTGGTGGCCGACCTTAGCBAACAGCAACT 1530
Db 181 ACAGATTGCCCTACCCCAAGTCTAGTGTGTGGTGGCCGACCTTAGCBAACAGCAACT 240

Qy 1531 GTCATGGCCATTGTCTACAAAGATTGCCCTACAGATGAACCTGAGAGGAGAGCTC 1590
Db 241 GTCATGGCCATTGTCTACAAAGATTGCCCTACAGATGAACCTGAGAGGAGAGCTC 300

Qy 1591 TGGAGGGTGGACATGCCCTCTGAAGCTCGTGATGATCGTTGGCATTCGATTGTACCATGAC 1650
Db 301 TGGAGGGTGGACATGCCCTCTGAAGCTCGTGATGATCGTTGGCATTCGATTGTACCATGAC 360

Qy 1651 ATGACAG 1657
Db 361 ATGACAG 367

```
RESULT 4
AAZ13404
ID AAZ13404 standard; cDNA; 300 BP.
XX
AC AAZ13404;
XX
DT 12-OCT-1999 (first entry)
DE Human gene expression product cDNA sequence SEQ ID NO:873.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN W09938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR -) CHIRON CORP.
PA (HYSE -) HYSEQ INC.
XX
PI Crkvenjakov R, Drmanac M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
XX WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
PS Claim 1; Page 862-863; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AAZ12532 to AAZ17779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell
XX suspected of being cancerous, where the gene product is encoded by one
XX of the 548 polynucleotide sequences given in AAZ12532 to AAZ1779. The
XX polynucleotides can be used as a source of primers and probes, which can
XX be used for a variety of purpose, e.g. detection of expression levels,
XX mapping, tissue typing or profiling, forensics, genetic analysis and
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX can be used for raising antibodies for experimental, diagnostic and
XX therapeutic purposes. The polynucleotides may also be used to construct
XX arrays for diagnostics (which may be used to determine function of an
XX encoded protein); and to detect differences in expression levels between
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX identify a genetic predisposition or susceptibility to a disease such as
XX cancer). The polynucleotides of the invention are especially used in the
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX and lung cancer. The polynucleotides can also be used to screen for
XX peptide analogues and antagonists.
XX
XX Sequence 300 BP; 75 A; 70 C; 91 G; 64 T; 0 other;
XX
XX Query Match 12.2%; Score 284; DB 20; Length 300;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-130;
XX Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1485 CCCAAGTCAGTGTGTGGTGGCCCGAACCTTAGGCAACAGCAACTGTGTCATGGCCATTGC 1544
XX
```

```
Db 1 CCCAAGTCAGTGTGTGTGGTGGCCCGAACCTTAGGCAACAGCAACTGTGTCATGGCCATTGC 60
Qy 1545 TACAAAGATTGCCCTACAGATGAAGCTGCAAGATGGAGGAGAGCTCTCTGGAGGTTGACAT 1604
Db 61 TACAAAGATTGCCCTACAGATGAAGCTGCAAGATGGAGGAGAGCTCTCTGGAGGTTGACAT 120
Qy 1605 CCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTTACCATGACATGACAGCTGGGCG 1664
Db 121 CCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTTTACCATGACATGACAGCTGGGCG 180
Qy 1665 GAGGTCAATCGCAGAGATTGTTGCCAGCATCAATGAAGGATGACCCCTGTTCTCAGG 1724
Db 181 GAGGTCAATCGCAGAGATTGTTGCCAGCATCAATGAAGGATGACCCCTGTTCTCAGG 240
Qy 1725 CTGCATATTTTCCAGGATAGAGGACAGGAGCTGCTAGATGGGCTCA 1768
Db 241 CTGCATATTTTCCAGGATAGAGGACAGGAGCTGCTAGATGGGCTCA 284

RESULT 5
AAF21880
ID AAF21880 standard; DNA; 423 BP.
XX
AC AAF21880;
XX
DT 27-MAR-2001 (first entry)
DE Human breast and ovarian cancer associated antigen gene SEQ ID 267.
XX
KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW neutropic; neurprotection; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antitumor; antiparasitic; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiac; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease; ds.
XX
OS Homo sapiens.
XX
FN W0200055173-A1.
XX
PD 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05881.
XX
PR 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-611515/S8.
XX
XX P-PSDB; AAB58977.
XX
XX New human breast and ovarian cancer associated gene sequences and the
XX polypeptides encoded by these genes, useful in the prevention,
XX treatment and diagnosis of cancer, immune disorders, cardiovascular
XX disorders and neurological diseases -
XX
XX Claim 1; Page 686; 1299pp; English.
XX
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
XX proteins AAB58711 - AAB59128. The DNA and protein sequences are
XX associated with breast and ovarian cancer. Included in the invention are
XX sequences AAF22032 - AAF22040 and AAB59129 which are used in the
XX isolation and characterisation of the DNA and protein sequences of the
XX invention. The breast and ovarian cancer associated DNA, protein, agonist
XX or antagonist sequences exhibit cytostatic; immunosuppressive;
XX neutropic; neurprotection; antiviral; antiallergic; hepatotropic;
XX antidiabetic; antiinflammatory; antitumor; antiparasitic; anticonvulsant;
XX antibacterial; antifungal; antiparasitic and cardiac activity. The
```

CC polynucleotide and protein sequences are used in the diagnosis of cancer,
CC particularly breast and ovarian cancer. The nucleic acid sequences,
CC proteins, agonists and agonists may also be used in the diagnosis,
CC prevention and treatment of immune disorders e.g. Addison's disease,
CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
CC arthritis and ulcerative colitis; cardiovascular disorders such as
CC myocardial ischaemias; wound healing; neurological diseases such as
CC cerebral anoxia and epilepsy; and infectious diseases.
XX
SQ Sequence 423 BP; 104 A; 98 C; 129 G; 88 T; 4 other;

Query Match 11.6%; Score 269; DB 21; Length 423;
Best Local Similarity 99.5%; Pred. No. 4.4e-123;
Matches 369; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1474 GATTGCCCTACCCCAAGTCAAGTGTGTGGTGGCCGACCTTAGGCAACAGCAAACTGTC 1533
Db |||||||
14 GATTGCCCTACCCCAAGTCAAGTGTGTGGTGGCCGACCTTAGGCAACAGCAAACTGTC 73

QY 1534 ATGGCCATTGCTACAAAGATTGCCCTACAGATGAATGCAAGATGGGAGGAGCTCTGG 1593
Db |||||||
74 ATGGCCATTGCTACAAAGATTGCCCTACAGATGAATGCAAGATGGGAGGAGCTCTGG 133

QY 1594 AGGTGACATCCCTCAAGCTCGTGCATGATCGTTGGCATGATGTTTACCATGACATG 1653
Db |||||||
134 AGGTGACATCCCTCAAGCTCGTGCATGATCGTTGGCATGATGTTTACCATGACATG 193

QY 1654 ACAGCTGGCGGAGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGATGACCCCG 1713
Db |||||||
194 ACAGCTGGCGGAGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGATGACCCCG 253

QY 1714 TGGTTCTCAGCTGCATATTTTCAGGATAGAGGACAGGAGCTGGTGGTCAAAAGTC 1773
Db |||||||
254 TGGTTCTCAGCTGCATATTTTCAGGATAGAGGACAGGAGCTGGTGGTCAAAAGTC 313

QY 1774 TGCCTGCAAGGGCTCTGAGGGCTTGAATAGCTGCAATGATGATACATGCCAGCCGGATC 1833
Db |||||||
314 TGCCTGCAAGGGCTCTGAGGGCTTGAATAGCTGCAATGATGATACATGCCAGCCGGATC 373

QY 1834 ATCGTGTAACCG 1844
Db |||||||
374 ATCGTGTAACCG 384

RESULT 6
ABZ20005
ID ABZ20005 standard; cDNA; 367 BP.
XX
AC ABZ20005;
XX
XX 23-JAN-2003 (first entry)
XX
XX Group III cDNA cancer related clone SEQ ID NO:2431.
XX
KW Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
KW immune response; virology; immunology; microbiology; molecular biology;
KW recombinant DNA technology; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200278516-A2.
XX
XX 10-OCT-2002.
XX
XX 28-MAR-2002; 2002WO-US10421.
XX
XX 30-MAR-2001; 2001US-280255P.
XX
XX 28-AUG-2001; 2001US-31563P.
XX
XX 09-JAN-2002; 2002US-347313P.
XX
XX (CORI-) CORIXA CORP.

PI Wang T, Wang S, Bangur CS, Gaiger A;
XX WPI; 2003-058387/05.
XX
XX New immunogenic polynucleotides or polypeptides useful for diagnosing,
XX preventing and treating cancer expressing CT or CP mRNA antigens, and
XX in virology, immunology, microbiology, molecular biology and
XX recombinant DNA techniques -
XX
XX Claim 1; SEQ ID 2431; 207pp; English.
XX
XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
XX ABP54446 to ABP54472 represent protein (II) sequences, from the present
XX invention. (I) and (II) have cytostatic activity and can be used in gene
XX therapy and vaccines. (I), (II), antibodies and compositions from the
XX present invention are useful for diagnosing, preventing and treating
XX cancer, which expresses CT or CP mRNA antigens. They are useful for
XX stimulating immune response. They can also be useful in virology,
XX immunology, microbiology, molecular biology and recombinant DNA
XX techniques.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 367 BP; 118 A; 78 C; 89 G; 79 T; 3 other;

Query Match 9.2%; Score 214; DB 25; Length 367;
Best Local Similarity 99.2%; Pred. No. 9.8e-96;
Matches 364; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1291 AAGTTTACACCGCATGGGCATGCAAAATGAGAAAAGCAATATGATTGAAGTGGATGAC 1350
Db 1 AAGTTTACACCGCATGGGCATGCAAAATGAGAAAAGCAATATGATTGAAGTGGATGAC 60

QY 1351 AGAATCTGAAGCCCTACTTAAGAGTCTTACAGCAAAAGGTCAAGCAGACACCCAGATAGTT 1410
Db 61 AGAATCTGAAGCCCTACTTAAGAGTCTTACAGCAAAAGGTCAAGCAGACACCCAGATAGTT 120

QY 1411 GTCTGTCTGTGTCAGTAATCGGAAGGACAAATACGATGCTTATTAATAAATACCTGTGT 1470
Db 121 GTCTGTCTGTGTCAGTAATCGGAAGGACAAATACGATGCTTATTAATAAATACCTGTGT 180

QY 1471 ACAGATTGCCCTACCCCAAGTCAAGTGTGGTGGCCCGACCTTAGGCAACAGCAAACT 1530
Db 181 ACAGATTGCCCTACCCCAAGTCAAGTGTGGTGGCCCGACCTTAGGCAACAGCAAACT 240

QY 1531 GTCATGCCCATTTGCTACAAAGATTGCCCTACAGATGAATGCAAGATGGGAGAGAGCTC 1590
Db 241 GTCATGCCCATTTGCTACAAAGATTGCCCTACAGATGAATGCAAGATGGGAGAGAGCTC 300

QY 1591 TGGAGGTGGACATCCCTCGAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGAC 1650
Db 301 TGGAGGTGGACATCCCTCGAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGAC 360

QY 1651 ATGACAG 1657
Db 361 ATGACAG 367

RESULT 7
ABZ19102
ID ABZ19102 standard; cDNA; 209 BP.
XX
XX ABZ19102;
XX
XX 23-JAN-2003 (first entry)
XX
XX Group III cDNA cancer related clone SEQ ID NO:1528.
XX
XX Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
XX immune response; virology; immunology; microbiology; molecular biology;
XX recombinant DNA technology; gene; ss.
XX


```
OS Homo sapiens.
XX WO200278516-A2.
XX 10-OCT-2002.
XX
XX 28-MAR-2002; 2002WO-US10421.
XX
XX 30-MAR-2001; 2001US-280255P.
XX 28-AUG-2001; 2001US-315563P.
XX 09-JAN-2002; 2002US-347313P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Wang S, Bangur CS, Gaiger A;
XX WPI; 2003-058387/05.
XX
XX New immunogenic polynucleotides or polypeptides useful for diagnosing,
XX preventing and treating cancer expressing CT or CP mRNA antigens, and
XX PT in virology, immunology, microbiology, molecular biology and
XX PT recombinant DNA techniques -
XX
XX Claim 1; SEQ ID 1528; 207pp; English.
XX
XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
XX CC ABP5446 to ABP5447 represent protein (II) sequences, from the present
XX CC invention. (I) and (II) have cytostatic activity and can be used in gene
XX CC therapy and vaccines. (I), (II), antibodies and compositions from the
XX CC present invention are useful for diagnosing, preventing and treating
XX CC cancer, which expresses CT or CP mRNA antigens. They are useful for
XX CC stimulating immune response. They can also be useful in virology,
XX CC immunology, microbiology, molecular biology and recombinant DNA
XX CC techniques.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 209 BP; 48 A; 47 C; 70 G; 44 T; 0 other;
SQ
Query Match 6.8%; Score 158; DB 25; Length 209;
Best Local Similarity 99.5%; Pred. No. 7e-68;
Matches 208; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1658 CTGGCGGAGTCAATCGCAGGATTGTTCAGCATCAATGAGGAGTACCGCGTGGT 1717
Db 1 CTGGCGGAGTCAATCGCAGGATTGTTCAGCATCAATGAGGAGTACCGCGTGGT 60
Qy 1718 TCTCACGCTGTATTTTCAGGATAGGACAGGAGCTGTAGATGGGCTCAAAGTCTGCC 1777
Db 61 TCTCACGCTGTATTTTCAGGATAGGACAGGAGCTGTAGATGGGCTCAAAGTCTGCC 120
Qy 1778 TGCAAGCGGCTCTGAGGCTTGGAAATAGCTCAATGATGATACATGCCCGCGATCATCG 1837
Db 121 TGCAAGCGGCTCTGAGGCTTGGAAATAGCTCAATGATGATGATGATGATGATGATGATG 180
Qy 1838 TGTACCGGATGGCGTAGGACAGCGCCAG 1866
Db 181 TGTACCGGATGGCGTAGGACAGCGCCAG 209
RESULT 8
ABN42800
ID ABN42800 standard; DNA; 60 BP.
XX
XX AC ABN42800;
XX
XX 15-JUL-2002 (first entry)
XX
XX Human spliced transcript detection oligonucleotide SEQ ID NO:1548.
XX
XX Human; mouse; rat; splice transcript; detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.
XX
```

```
XX Homo sapiens.
XX OS WO200210449-A2.
XX PN 07-FEB-2002.
XX PD
XX PF 20-JUL-2001; 2001WO-IB01903.
XX PF
XX 28-JUL-2000; 2000US-221607P.
XX PR 02-MAY-2001; 2001US-287724P.
XX PR
XX (COMP-) COMPUGEN INC.
XX PA
XX Shoshan A, Wasserman A, Mintz E, Mintz L, Paigler S;
XX WPI; 2002-257383/30.
XX
XX New oligonucleotide libraries comprising oligonucleotides which
XX PT selectively hybridize to mRNAs transcribed from a transcription unit of
XX PT a genome, useful for detecting tissue-, pathology-, and
XX PT developmental-specific genes -
XX
XX Example 1; SEQ ID 15548; 47pp; English.
XX
XX The present invention describes oligonucleotide libraries for detecting
XX CC messenger RNAs that populate a (sub-)transcriptome, where the
XX CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX CC transcription units that populate a genome. The library comprises
XX CC several oligonucleotides, each capable of hybridising selectively to a
XX CC set of messenger RNAs transcribed from a given transcription unit of
XX CC the genome, which encodes one or more messenger RNA splice variants.
XX CC The oligonucleotide libraries are useful for detecting mRNAs from a
XX CC biological sample, in expression profiling studies, in qualitatively or
XX CC quantitatively characterising the corresponding transcriptome, and in
XX CC detecting RNA transcripts and splice variants of human or animal
XX CC transcriptomes. The libraries may also be used as specialised mini-
XX CC libraries to detect transcripts of a sub-transcriptome under a
XX CC particular biological or pathological state, and so allowing the
XX CC detection of tissue- and pathology-specific genes such as those genes
XX CC only expressed in specific tissue under a specific pathological
XX CC condition; to detect developmental specific genes; and to detect RNA
XX CC transcripts and splice variants of a transcriptome of a patient suffering
XX CC from a particular disorder. ABN27253 to ABN59589 represent
XX CC oligonucleotide sequences from rats, humans and mice, which are used in
XX CC the exemplification of the present invention.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 60 BP; 23 A; 9 C; 17 G; 11 T; 0 other;
SQ
Query Match 2.6%; Score 60; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. No. 3.8e-19;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1924 GGTAGAGTTTACACCTAGACTAAGCGTAATCTGTGTGAGAAAGAGTGAAACACGAGA 1983
Db 1 GGTAGAGTTTACACCTAGACTAAGCGTAATCTGTGTGAGAAAGAGTGAAACACGAGA 60
RESULT 9
AAS25525/C
ID AAS25525 standard; cDNA; 460 BP.
XX
XX AC AAS25525;
XX
XX 07-NOV-2001 (first entry)
XX
XX Human ovarian PCR-subtracted cDNA library clone #1610.
XX
XX Immunogenic protein; cancer; ovarian tumour; T-cell stimulation; ss;
XX gene therapy; cytostatic; T-cell expansion; nucleic acid hybridisation;
XX
```


PT New colon cancer polypeptides and polynucleotides, useful as vaccines,
PT for diagnosing, preventing, and treating colon cancer, and as markers
PT for the progression of cancer -
XX
PS Claim 1; SEQ ID No 34; 147pp; English.
XX
CC The invention relates to polynucleotides encoding colon tumour proteins.
CC The polynucleotides and encoded polypeptides are useful in pharmaceutical
CC compositions, such as vaccines, for the diagnosis, prevention, and
CC treatment of colon cancer. Polynucleotide sequences may be used as
CC hybridisation probes or primers, and in the design and preparation of
CC ribozyme molecules for inhibiting expression of tumour polypeptides and
CC proteins in tumour cells. The compositions are useful for stimulating an
CC immune response against cancer, particularly for the immunotherapy of
CC colon cancer, and as markers for the progression of cancer.
CC ABK4450-ABK46237 represent coding sequences of human colon tumour
CC proteins of the invention.
CC Note: With the exception of SEQ ID No 1 and 2, the sequence data
CC for this patent did not form part of the printed specification but was
CC supplied by the European Patent Office.
XX
SQ Sequence 501 BP; 134 A; 81 C; 93 G; 193 T; 0 other;
Query Match 1.5%; Score 34; DB 24; Length 501;
Best Local Similarity 100.0%; Pred. No. 3.2e-06;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2271 CCAGATATTTCAGAGAGGCCAAATCTGTCACTG 2304
Db 1 CCAGATATTTCAGAGAGGCCAAATCTGTCACTG 34
RESULT 12
ABX15293/C
ID ABX15293 standard; DNA; 22 BP.
XX
AC ABX15293;
XX
DT 27-MAR-2003 (first entry)
XX
DE ASM 698-13 B_1 PCR primer ASM75-3' T.
XX
KW ss: allele frequency; proofreading polymerase; gene typing; PCR;
KW karyotyping; genotyping; DNA family planning; diagnostics; primer;
KW prenatal testing; paternal determination; pharmacogenetics;
KW forensic analysis; ASM 698-13 B_1.
XX
OS Unidentified.
XX
PH Key Location/Qualifiers
FT modified_base 22 /*tag= a
FT /*mod_base= OTHER
FT /note= "T is labeled with a ROX (not defined) moiety"
XX
US2002142336-A1.
XX
PD 03-OCT-2002.
XX
XX 01-FEB-2002; 2002US-0061961.
XX
XX 02-FEB-2001; 2001US-266035P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Smith DR, Thomann H, Cahill P;
XX
XX WPI; 2003-182355/18.
XX
XX Detecting the presence or absence of a first nucleotide at position
PT within a strand of DNA, useful in gene typing, genotyping, disease
PT diagnostics, prenatal testing, paternal determination, pharmacogenetics
PT and forensic analysis -

XX
PS Example 2; Page 13; 40pp; English.
XX
CC The invention relates to detecting the presence or absence of a first
CC nucleotide, at a position within a strand of DNA, comprising forming an
CC admixture of a primer and the strand of DNA, with or without a mixture of
CC labeled dideoxynucleotides and forming a hybridisation product, where the
CC primer comprises an electrophoretic tag (e-tag) group or fluorescent
CC label opposite the first nucleotide. Also included is a method for
CC determining allele frequency at a first nucleotide position within a
CC strand of DNA in a sample comprising: (a) providing a first primer
CC comprising a sequence of DNA which hybridises with the strand of DNA
CC adjacent to the first nucleotide position, and having a second nucleotide
CC opposite the first nucleotide position, the second nucleotide hybridising to the
CC with a first detectable label, the second nucleotide hybridising to the
CC first nucleotide in the event the second nucleotide not hybridising to the
CC first nucleotide in the event the second nucleotide is not complementary;
CC (b) providing a second primer, the second primer comprising a sequence
CC of DNA which hybridises with the strand of DNA adjacent to the first
CC nucleotide position, and having a third nucleotide opposite the first
CC nucleotide position, the third nucleotide associated with a second
CC detectable label, and hybridising or not hybridising to the first
CC complementary to the first nucleotide; (c) forming an admixture of the
CC first and second primers and the strand of DNA to form a
CC hybridisation product; (d) applying a proofreading polymerase to the
CC hybridisation product under conditions in which the second and third
CC nucleotide is preferentially excised in the event the second and third
CC nucleotide is not hybridised to the first nucleotide, and in which the
CC second and third nucleotide is preferentially incorporated into an
CC extension product in the event the second and third nucleotide is
CC hybridised to the first nucleotide; and (e) monitoring the sample for the
CC presence of a first or a second label in association with the extension
CC product, where the ratio of the first and second label is indicative of
CC allele frequency at the first nucleotide position within a strand of DNA
CC in a sample. The methods of the present invention are useful in gene
CC typing, karyotyping, genotyping, DNA family planning, diagnostics,
CC prenatal testing, paternal determination, pharmacogenetics and forensic
CC analysis. The present sequence is a labeled PCR primer used to
CC investigate the ASM 698-13 B_1 (not defined) gene region using the method
CC of the invention.
XX
SQ Sequence 22 BP; 5 A; 5 C; 8 G; 4 T; 0 other;
Query Match 0.9%; Score 21; DB 25; Length 22;
Best Local Similarity 100.0%; Pred. No. 9.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 871 CCAGGGCGCTGCCATGCTCATT 891
Db 21 CCAGGGCGCTGCCATGCTCATT 1
RESULT 13
AAV21209
ID AAV21209 standard; DNA; 1664976 BP.
XX
AC AAV21209;
XX
DT 10-NOV-1998 (first entry)
XX
DE Methanococcus jannaschii circular chromosome.
XX
KW Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KW genome; autotrophic; extrachromosomal element; identification; ds.
XX
OS Methanococcus jannaschii.
XX
XX WO9807830-A2.
XX
PD 26-FEB-1998.
XX

```

PF 22-AUG-1997; 97WO-US14900.
XX
XX
XX 22-AUG-1996; 96US-0024428.
XX
XX (GENO-) INST GENOMIC RES.
XX (UNII) UNIV ILLINOIS FOUND.
XX (UJJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
XX Bult CJ, Smith HO, Venter JC, White OR, Woese CR;
XX WPI; 1998-169145/15.
XX
XX Complete genome sequence of methano-genic archaeon, Methanococcus
XX Jannaschii - useful in identification of M. jannaschii genome
XX fragment
XX
XX Claim 13; Page 152-585; 614pp; English.
XX
XX The present sequence represents the complete 1.66-megabase pair genome
XX sequence of the Methanococcus Jannaschii circular chromosome. The
XX present invention describes M. jannaschii open reading frames from the
XX genome sequence. The invention also describes a computer based system
XX for identifying fragments of the M. jannaschii genome that are
XX homologous to target nucleotide sequences, comprising: (a) data storage
XX means comprising the nucleotide sequence of the 1664976, 58407 or 16550
XX bp sequence (see AAV21209, AAV21210 and AAV21211), or a nucleotide
XX sequence at least 99.9% identical to it; (b) search means for comparing a
XX target sequence to the nucleotide sequence of the data storage means to
XX identify a homologous sequence, and (c) retrieval means for obtaining
XX the homologous sequence. The method, which is based on whole genome
XX random sequencing of an autotrophic archaeon M. jannaschii, the genome
XX of which consists of 3 physically distinct elements, a large circular
XX chromosome (the 1664976 bp sequence given in AAV21209), a large circular
XX extra-chromosomal element (the 58407 bp sequence given in AAV21210), and
XX a small circular extra-chromosomal element (the 16550 bp sequence given
XX in AAV21211), can be used in the identification of M. jannaschii genome
XX fragment.
XX
XX Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573992 T; 101 other;
XX
XX Query Match 0.9%; Score 21; DB 19; Length 1664976;
XX Best Local Similarity 100.0%; Pred. No. 8.5;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 AGTTCTATAATATTATTTC 391
Db 430031 AGTTCTATAATATTATTTC 430051

RESULT 14
ABX15292/c
ID ABX15292 standard; DNA; 19 BP.
XX
XX AC ABX15292;
XX
XX DT 27-MAR-2003 (first entry)
XX
XX DE ASM 698-13 B_1 PCR primer ASM73-3' C.
XX
XX ss; allele frequency; proofreading polymerase; gene typing; PCR;
XX karyotyping; genotyping; DNA family planning; diagnostics; primer;
XX prenatal testing; paternal determination; pharmacogenetics;
XX forensic analysis; ASM 698-13 B_1.
XX
XX Unidentified.
XX
XX Key Location/Qualifiers
XX modified_base 19
XX /*tag= a
XX /mod_base= OTHER
XX /note= "C is labeled with a 6-carbo fluorescein
XX (FAM) moiety"
XX

```

```

PN US2002142336-A1.
XX
XX PD 03-OCT-2002.
XX
XX PF 01-FEB-2002; 2002US-0061961.
XX
XX PR 02-FEB-2001; 2001US-266035P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Smith DR, Thomann H, Cahill P;
XX
XX WPI; 2003-182355/18.
XX
XX Detecting the presence or absence of a first nucleotide at position
XX within a strand of DNA, useful in gene typing, genotyping, disease
XX diagnostics, prenatal testing, paternal determination, pharmacogenetics
XX and forensic analysis -
XX
XX Example 2; Page 13; 40pp; English.
XX
XX The invention relates to detecting the presence or absence of a first
XX nucleotide, at a position within a strand of DNA, comprising forming an
XX admixture of a primer and the strand of DNA, with or without a mixture of
XX labeled dideoxynucleotides and forming a hybridisation product, where the
XX primer comprises an electrophoretic tag (e-tag) group or fluorescent
XX label opposite the first nucleotide. Also included is a method for
XX determining allele frequency at a first nucleotide position within a
XX strand of DNA in a sample comprising: (a) providing a first primer
XX comprising a sequence of DNA which hybridises with the strand of DNA
XX adjacent to the first nucleotide position, and having a second nucleotide
XX opposite the first nucleotide position, the second nucleotide associated
XX with a first detectable label, the second nucleotide hybridising to the
XX first nucleotide in the event the second nucleotide is complementary to
XX the first nucleotide and the second nucleotide not hybridising to the
XX first nucleotide in the event the second nucleotide is not complementary;
XX (b) providing a second primer, the second primer comprising a sequence
XX of DNA which hybridises with the strand of DNA adjacent to the first
XX nucleotide position, and having a third nucleotide opposite the first
XX nucleotide position, the third nucleotide associated with a second
XX detectable label, and hybridising or not hybridising to the first
XX nucleotide in the event the third nucleotide is complementary or not
XX complementary to the first nucleotide; (c) forming an admixture of the
XX first and second primers and the strand of DNA to form a
XX hybridisation product; (d) applying a proofreading polymerase to the
XX hybridisation product under conditions in which the second and third
XX nucleotide is preferentially excised in the event the second and third
XX nucleotide is not hybridised to the first nucleotide, and in which the
XX second and third nucleotide is preferentially incorporated into an
XX extension product in the event the second and third nucleotide is
XX hybridised to the first nucleotide; and (e) monitoring the sample for the
XX presence of a first or a second label in association with the extension
XX product, where the ratio of the first and second label is indicative of
XX allele frequency at the first nucleotide position within a strand of DNA
XX in a sample. The methods of the present invention are useful in gene
XX typing, karyotyping, genotyping, DNA family planning, diagnostics,
XX prenatal testing, paternal determination, pharmacogenetics and forensic
XX analysis. The present sequence is a labeled PCR primer used to
XX investigate the ASM 698-13 B_1 (not defined) gene region using the method
XX of the invention.
XX
XX Sequence 19 BP; 3 A; 6 C; 8 G; 2 T; 0 other;
XX
XX Query Match 0.8%; Score 19; DB 25; Length 19;
XX Best Local Similarity 100.0%; Pred. No. 95;
XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 870 GCCAGGCGCTGCATGCTC 888
Db 19 GCCAGGCGCTGCATGCTC 1

RESULT 15

```

ABN23657/c
ID ABN23657 standard; cDNA; 302 BP.
XX AC ABN23657;
XX DT 24-JUN-2002 (first entry)
XX DE Human ORFX polynucleotide sequence SEQ ID NO:15791.
XX KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis; Gene; ss.
XX OS Homo sapiens.
XX PN WO200192523-A2.
XX PD 06-DEC-2001.
XX PF 29-MAY-2001; 2001WO-US10836.
XX PR 30-MAY-2000; 2000US-206132P.
XX PR 29-AUG-2000; 2000US-228716P.
XX PA (CURA-) CURAGEN CORP.
XX PI Shimkets RA, Leach MD;
XX DR WPI; 2002-106308/14.
XX DR P-PSDB; ABP07905.

QY 635 AAGTTTCCAAAGAACTAAT 653
Db 165 AAGTTTCCAAAGAACTAAT 147

Search completed: December 6, 2003, 18:22:26
Job time : 622 secs

Novel human polypeptides and polynucleotides useful for diagnosing, preventing and treating cardiovascular disease, neurodegenerative, hyperproliferative disorders and autoimmune disorders -
Disclosure; SEQ ID 15791; 1037pp; English.
The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1 in the specification) ABN15762 to ABN27252 encode the human ORFX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, or periodontal disease, and for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues and conditions resulting from systemic cytokine damage.
N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 302 BP; 90 A; 54 C; 58 G; 96 T; 4 other;

Query Match 0.8%; Score 19; DB 24; Length 302;
Best Local Similarity 100.0%; Pred.No. 92;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: December 10, 2003, 17:32:26 ; Search time 42 Seconds
(without alignments)
4761.685 Million cell updates/sec

Title: US-10-043-774B-2
Perfect score: 775
Sequence: 1 MIFGNTFQNLHDHYKESKTG.....VQGSTHREPNSLNRLYYL 775

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL.23.*

1: sp archea.*
2: sp bacteria.*
3: sp fungi.*
4: sp human.*
5: sp invertebrate.*
6: sp mammal.*
7: sp mhc.*
8: sp organelle.*
9: sp phage.*
10: sp plant.*
11: sp rodent.*
12: sp virus.*
13: sp vertebrate.*
14: sp unclassified.*
15: sp rvirus.*
16: sp bacteriaph.*
17: sp archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	775	100.0	775	Q96JD5	Q96jd5 homo sapien
2	772	99.6	861	Q96J94	Q96j94 homo sapien
3	645	83.2	829	Q8NA60	Q8na60 homo sapien
4	570	73.5	861	Q95404	Q95404 homo sapien
5	469	60.5	861	Q8TBY5	Q8tby5 homo sapien
6	123	15.9	862	Q9JMB7	Q9jmb7 mus musculu
7	21	2.7	858	Q8UVX0	Q8uvx0 brachydanio
8	16	2.1	808	Q9GPA8	Q9gpa8 strongyloce
9	16	2.1	854	Q9GPA7	Q9gpa7 strongyloce
10	15	1.9	722	Q17567	Q17567 caenorhabdi
11	15	1.9	824	P90786	P90786 caenorhabdi
12	11	1.4	121	Q8CC75	Q8cc75 mus musculu
13	11	1.4	666	Q8N9V8	Q8n9v8 homo sapien
14	11	1.4	852	Q8N9V8	Q8n9v8 homo sapien
15	11	1.4	852	Q8N9H2	Q8n9h2 homo sapien
16	11	1.4	878	Q8CGT6	Q8cgt6 mus musculu

17	9	1.2	189	5	Q95P85	Q95pe5 drosophila
18	9	1.2	866	5	O76922	O76922 drosophila
19	9	1.2	1035	5	O21079	O21079 caenorhabdi
20	9	1.2	1037	5	Q9TW94	Q9tw94 caenorhabdi
21	8	1.0	371	4	Q9NW28	Q9nw28 homo sapien
22	8	1.0	398	10	O80872	O80872 arabidopsis
23	8	1.0	421	5	Q9GPA6	Q9gpa6 strongyloce
24	8	1.0	449	10	O8S9X7	O8s9x7 oryza sativ
25	8	1.0	530	4	Q96SW6	Q96sw6 homo sapien
26	8	1.0	580	11	Q99MV6	Q99mv6 mus musculu
27	8	1.0	704	5	Q95XQ7	Q95xq7 caenorhabdi
28	8	1.0	730	2	O07667	O07667 enterococu
29	8	1.0	971	11	Q9JMB6	Q9jmb6 mus musculu
30	8	1.0	971	11	O8CDG1	O8cdg1 mus musculu
31	8	1.0	973	4	O8TCS9	O8tcs9 homo sapien
32	8	1.0	1137	5	Q93250	Q93250 caenorhabdi
33	8	1.0	1498	13	Q8UUM8	Q8uum8 oryzias lat
34	7	0.9	41	2	O8G96	O8g96 neisseria p
35	7	0.9	61	16	O8YTK9	O8ytk9 anabaena sp
36	7	0.9	68	16	O8FKT9	O8fkt9 escherichia
37	7	0.9	77	16	O9ZJ23	O9zj23 helicobacte
38	7	0.9	83	17	O8TSK7	O8tsk7 methanosarc
39	7	0.9	85	2	Q9S556	Q9s556 pseudomonas
40	7	0.9	85	16	Q9S564	Q9s564 pseudomonas
41	7	0.9	86	8	Q95817	Q95817 hymenolepis
42	7	0.9	86	16	O8PDT6	O8pdt6 escherichia
43	7	0.9	90	2	O8GC58	O8gc58 escherichia
44	7	0.9	92	2	Q93F09	Q93f09 shigella fl
45	7	0.9	92	16	O8FG80	O8fg80 escherichia

ALIGNMENTS

RESULT 1
Q96JD5 PRELIMINARY; PRT; 775 AA.
ID Q96JD5; AC Q96JD5; DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE HIWI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Sharma A.K., Nelson M.C., Brandt J.E., Wessman M., Muhmud N.,
RA Weller K.P., Hoffman R.;
RT "Human CD34+ Stem Cells Express the hiwi Gene, a Human Homolog of the
RT Drosophila Gene piwi.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AF264004; AAK92281.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 775 AA; 89484 MW; DF169A2E9EAFD916 CRC64;

Query Match 100.0%; Score 775; DB 4; Length 775;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MIFGNTFQNLHDHYKESKTGSGGIIVRLSTHFRITSPQWALYQYHIDYNPLMEARRLR 60
DB 1 MIFGNTFQNLHDHYKESKTGSGGIIVRLSTHFRITSPQWALYQYHIDYNPLMEARRLR 60
QY 61 SALLFQHEDLICKCHAFDGTILFLPKRQQKQVTEFSKTRNGEDVRIITITLNEPLPTSP 120

Db 61 SALLFQHEDLIGKCHAFDGTILFLPKRLQKQYEVFSKTRNGEDVRITITLTNELPPTSP 120
 QY 121 TCIQFYNIIFRLLKIMNLQOIGRNYNPNPDIDPSHRLVIWPGFTTTSILOYENSIMLC 180
 Db 121 TCIQFYNIIFRLLKIMNLQOIGRNYNPNPDIDPSHRLVIWPGFTTTSILOYENSIMLC 180
 QY 181 TDVSHKVLASETVLDPMFNFYHOTEEHKFQEQVSKELIGLVLTCKYNNKTYVDDIDWDQ 240
 Db 181 TDVSHKVLASETVLDPMFNFYHOTEEHKFQEQVSKELIGLVLTCKYNNKTYVDDIDWDQ 240
 QY 241 NPKSTFKKADGSSVSFLYYRKYNOEITDLKQPVLSQPKRRRGGTLPGAMLIPEL 300
 Db 241 NPKSTFKKADGSSVSFLYYRKYNOEITDLKQPVLSQPKRRRGGTLPGAMLIPEL 300
 QY 301 CYLTGLTDMRNDFNVMKDLAVHTRLTPEQREVGLLDYIHKNDNVQRELDMGLSPD 360
 Db 301 CYLTGLTDMRNDFNVMKDLAVHTRLTPEQREVGLLDYIHKNDNVQRELDMGLSPD 360
 QY 361 SNLLSFGRILOTEKHOGGKTDPYNPOFADWSKETRGAPLISVKPLDNWLLIYTERNYE 420
 Db 361 SNLLSFGRILOTEKHOGGKTDPYNPOFADWSKETRGAPLISVKPLDNWLLIYTERNYE 420
 QY 421 AANSLIQLFKVTPAMGMQMKAIMIEVDRTTEAYLRVLQKVTADTQIVVCLLSNRKD 480
 Db 421 AANSLIQLFKVTPAMGMQMKAIMIEVDRTTEAYLRVLQKVTADTQIVVCLLSNRKD 480
 QY 481 KYDAIKKYLCTDPTSPQCVVARTLGKQQTVAIAIKIALQMNCKMGELWRVDPILKLV 540
 Db 481 KYDAIKKYLCTDPTSPQCVVARTLGKQQTVAIAIKIALQMNCKMGELWRVDPILKLV 540
 QY 541 MIYVIGDCHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAW 600
 Db 541 MIYVIGDCHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAW 600
 QY 601 NSCNEMPSRIIYVRDGVGGQKLTIVNVEVPQFLDCLKSGRGYNPRLTVIVVKRVNT 660
 Db 601 NSCNEMPSRIIYVRDGVGGQKLTIVNVEVPQFLDCLKSGRGYNPRLTVIVVKRVNT 660
 QY 661 RPPAQSGRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSVSGSPHYNVYDNSGLKPD 720
 Db 661 RPPAQSGRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSVSGSPHYNVYDNSGLKPD 720
 QY 721 HIQRLTYKLCHIIYNNPVGIRVPAPCOYAKHLAFLVGSQSIHREPNSLSNRLYYL 775
 Db 721 HIQRLTYKLCHIIYNNPVGIRVPAPCOYAKHLAFLVGSQSIHREPNSLSNRLYYL 775

RESULT 2

Q96J94 PRELIMINARY; PRT; 861 AA.
 AC Q96J94;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE PIWI protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Sha J.H.;
 RT "Cloning and identification of human piwi protein related to testis development."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF387507; AAK69348.1; --
 DR InterPro; IPR003100; PAZ.
 DR InterPro; IPR003165; PIWI.
 DR Pfam; PF02170; PAZ; 1.
 DR Pfam; PF02171; Piwi; 1.
 DR PROSITE; PS50821; PAZ; 1.
 DR PROSITE; PS50822; PIWI; 1.

SQ SEQUENCE 861 AA; 98603 MW; 58D7P6C7321DEFA4 CRC64;
 Query Match 99.6%; Score 772; DB 4; Length 861;
 Best Local Similarity 100.0%; Pred.No. 0;
 Matches 772; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 GVNTRQNLDHVKESKTSGGSIIVRLSTNNHPLTTSRPOWALYQXHIDYNPLMEARLRSL 63
 Db 90 GVNTRQNLDHVKESKTSGGSIIVRLSTNNHPLTTSRPOWALYQXHIDYNPLMEARLRSL 149
 QY 64 LFQHEDLIGKCHAFDGTILFLPKRLQKQYEVFSKTRNGEDVRITITLTNELPPTSPCL 123
 Db 150 LFQHEDLIGKCHAFDGTILFLPKRLQKQYEVFSKTRNGEDVRITITLTNELPPTSPCL 209
 QY 124 QFYNIIFRLLKIMNLQOIGRNYNPNPDIDPSHRLVIWPGFTTTSILOYENSIMLCITDV 183
 Db 210 QFYNIIFRLLKIMNLQOIGRNYNPNPDIDPSHRLVIWPGFTTTSILOYENSIMLCITDV 269
 QY 184 SHKVLRSSTVLDPMFNFYHOTEEHKFQEQVSKELIGLVLTCKYNNKTYVDDIDWDQNP 243
 Db 270 SHKVLRSSTVLDPMFNFYHOTEEHKFQEQVSKELIGLVLTCKYNNKTYVDDIDWDQNP 329
 QY 244 STFKKADGSSVSFLYYRKYNOEITDLKQPVLSQPKRRRGGTLPGAMLIPELCYL 303
 Db 330 STFKKADGSSVSFLYYRKYNOEITDLKQPVLSQPKRRRGGTLPGAMLIPELCYL 389
 QY 304 TGLTDMRNDFNVMKDLAVHTRLTPEQREVGLLDYIHKNDNVQRELDMGLSPD 363
 Db 390 TGLTDMRNDFNVMKDLAVHTRLTPEQREVGLLDYIHKNDNVQRELDMGLSPD 449
 QY 364 LSFSGRILQTEKHOGGKTDPYNPOFADWSKETRGAPLISVKPLDNWLLIYTERNYEAA 423
 Db 450 LSFSGRILQTEKHOGGKTDPYNPOFADWSKETRGAPLISVKPLDNWLLIYTERNYEAA 509
 QY 424 SLIQNLFKVTPAMGMQMKAIMIEVDRTTEAYLRVLQKVTADTQIVVCLLSNRKDYD 483
 Db 510 SLIQNLFKVTPAMGMQMKAIMIEVDRTTEAYLRVLQKVTADTQIVVCLLSNRKDYD 569
 QY 484 AIKKYLCTDPTSPQCVVARTLGKQQTVAIAIKIALQMNCKMGELWRVDPILKLV 543
 Db 570 AIKKYLCTDPTSPQCVVARTLGKQQTVAIAIKIALQMNCKMGELWRVDPILKLV 629
 QY 544 GIDCVHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAWNSC 603
 Db 630 GIDCVHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAWNSC 689
 QY 604 NEYMPRSRIIYVRDGVGGQKLTIVNVEVPQFLDCLKSGRGYNPRLTVIVVKRVNTRFF 663
 Db 690 NEYMPRSRIIYVRDGVGGQKLTIVNVEVPQFLDCLKSGRGYNPRLTVIVVKRVNTRFF 749
 QY 664 AQSGRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSVSGSPHYNVYDNSGLKPDHIQ 723
 Db 750 AQSGRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSVSGSPHYNVYDNSGLKPDHIQ 809
 QY 724 RLTYKLCHIIYNNPVGIRVPAPCOYAKHLAFLVGSQSIHREPNSLSNRLYYL 775
 Db 810 RLTYKLCHIIYNNPVGIRVPAPCOYAKHLAFLVGSQSIHREPNSLSNRLYYL 861

RESULT 3

Q8NA60 PRELIMINARY; PRT; 829 AA.
 ID Q8NA60
 AC Q8NA60;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein FLJ35814 (piwi).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

```
RC TISSUE=Testis;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK093133; BAC04068.1; -.
DR InterPro; IPR0031100; PAZ.
DR InterPro; IPR0031165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
KW Hypothetical protein.
KW
SQ SEQUENCE 829 AA, 94816 MW, 3AF9FD1D0FA9F5A7 CRC64;
Query Match 83.2%; Score 645; DB 4; Length 829;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 645; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 93 TEVSKTRNGEDVRITITLTNELPPTSPCLQFYNIIPRLKIMNLQOIGRNYNPNDDP 152
DB 179 TEVSKTRNGEDVRITITLTNELPPTSPCLQFYNIIPRLKIMNLQOIGRNYNPNDDP 238
QY 153 IDIPSHRLVWPGFTTSLIQENSIMLCTDVSHKVLRSVLDPMFNFYHOTEHKKFQEQ 212
DB 239 IDIPSHRLVWPGFTTSLIQENSIMLCTDVSHKVLRSVLDPMFNFYHOTEHKKFQEQ 298
QY 213 VSKELIGLVLTAKYNNKTVRVDIDWDQNPSTFKKADSEVSFLEYKQYNQIETDLK 272
DB 299 VSKELIGLVLTAKYNNKTVRVDIDWDQNPSTFKKADSEVSFLEYKQYNQIETDLK 358
QY 273 QPVLVSQPKRRPGGTLPGPAMLIPELCVLTGLTDKMDNFVMDKLVHTRLTPEQEQ 332
DB 359 QPVLVSQPKRRPGGTLPGPAMLIPELCVLTGLTDKMDNFVMDKLVHTRLTPEQEQ 418
QY 333 REVGRLLIDYIHKNDNVQRELDMGLSFDNSLLSFGSRILQTEKHOGGKTFDYNPOFADW 392
DB 419 REVGRLLIDYIHKNDNVQRELDMGLSFDNSLLSFGSRILQTEKHOGGKTFDYNPOFADW 478
QY 393 SKETRGAPLISVKPLDNWLLIYTRRNYEANSLLI QNLKVTTPANGMQRKAIMIEVDRT 452
DB 479 SKETRGAPLISVKPLDNWLLIYTRRNYEANSLLI QNLKVTTPANGMQRKAIMIEVDRT 538
QY 453 EAYLRVLOQKVATDTQIVVCLSSNRKDYDAIKKYLCTDCPTPSQCVVARTLGKQOTVM 512
DB 539 EAYLRVLOQKVATDTQIVVCLSSNRKDYDAIKKYLCTDCPTPSQCVVARTLGKQOTVM 598
QY 513 AIATKIALQMNCKMGGLWRVDIPLKLVIMVIGDCYHDMTAGRRSIAGFVASINEGTRW 572
DB 599 AIATKIALQMNCKMGGLWRVDIPLKLVIMVIGDCYHDMTAGRRSIAGFVASINEGTRW 658
QY 573 FSCCIFODRQQLVDGLKVCLOALRAWNSCNEVPSRIIIVYRDGVDGGLKTLVNYEVP 632
DB 659 FSCCIFODRQQLVDGLKVCLOALRAWNSCNEVPSRIIIVYRDGVDGGLKTLVNYEVP 718
QY 633 QFLDCLKSGRGYNPRLTIVVKKRVNTRFFAQSGLQNPGLPCTVIDVETRPEDWDF 692
DB 719 QFLDCLKSGRGYNPRLTIVVKKRVNTRFFAQSGLQNPGLPCTVIDVETRPEDWDF 778
QY 693 IVSQAVRSGSVSPHYNYVINDNSGLKPDHILQRLTYKLCHIIYNNWP 737
DB 779 IVSQAVRSGSVSPHYNYVINDNSGLKPDHILQRLTYKLCHIIYNNWP 823
RESULT 4
O95404 PRELIMINARY; PRT; 861 AA.
ID O95404
AC O95404;
```

```
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE HIWI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99069219; PubMed=9851978;
RA Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H.;
RT "A novel class of evolutionarily conserved genes defined by piwi are
essential for stem cell self-renewal.";
RL Genes Dev. 12:3715-3727(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Qiao D., Zeeman A.-M., Deng W., Looijenga L.H.J., Lin H.;
RT "Molecular characterization of hiwi, a human member of the piwi stem
cell gene family whose overexpression is correlated to seminomas.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF104260; AAC97371.2; -.
DR Genew; HGNC:9007; PIWI1.
DR InterPro; IPR0031100; PAZ.
DR InterPro; IPR0031165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 861 AA, 98530 MW, 43D7F60E99D997B7 CRC64;
Query Match 73.5%; Score 570; DB 4; Length 861;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 770; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 GVNTRQNLDHVKESKSGSIIVRLSTNHFRLTSRQWALYQYHIDYNPLMEARLSAL 63
DB 90 GVNTRQNLDHVKESKSGSIIVRLSTNHFRLTSRQWALYQYHIDYNPLMEARLSAL 149
QY 64 LFOHEDLIGKCHAFGTILFLPKLQKKVTEVFSKTRNGEDVRITITLTNELPPTSPCL 123
DB 150 LFOHEDLIGKCHAFGTILFLPKLQKKVTEVFSKTRNGEDVRITITLTNELPPTSPCL 209
QY 124 QFYNIIFRLLKIMNLQOIGRNYNPNDDIPSHRLVWPGFTTSLIQENSIMLCTDV 183
DB 210 QFYNIIFRLLKIMNLQOIGRNYNPNDDIPSHRLVWPGFTTSLIQENSIMLCTDV 269
QY 184 SHKVLRSVLDPMFNFYHOTEHKKFQEQVSKELIGLVLTAKYNNKTVRVDIDWDQNP 243
DB 270 SHKVLRSVLDPMFNFYHOTEHKKFQEQVSKELIGLVLTAKYNNKTVRVDIDWDQNP 329
QY 244 STFKKADGSEVSFLEYKQYNQIETDLKQPVLSQPKRRPGGTLPGPAMLIPELCYL 303
DB 330 STFKKADGSEVSFLEYKQYNQIETDLKQPVLSQPKRRPGGTLPGPAMLIPELCYL 389
QY 304 TGLTDKMDNFVMDKLVHTRLTPEQOREVGRLLIDYIHKNDNVQRELDMGLSFDNSL 363
DB 390 TGLTDKMDNFVMDKLVHTRLTPEQOREVGRLLIDYIHKNDNVQRELDMGLSFDNSL 449
QY 364 LQFSGRILQTEKHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA 423
DB 450 LQFSGRILQTEKHOGGKTFDYNPOFADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA 509
QY 424 SLIQNLKFKVTPANGMQRKAIMIEVDRTPEAYLRVLOQKVATDTQIVVCLSSNRKDY 483
DB 510 SLIQNLKFKVTPANGMQRKAIMIEVDRTPEAYLRVLOQKVATDTQIVVCLSSNRKDY 569
QY 484 AIKKYLCTDCPTPSQCVVARTLGKQOTVMAIATKIALQMNCKMGGLWRVDIPLKLVIM 543
DB 570 AIKKYLCTDCPTPSQCVVARTLGKQOTVMAIATKIALQMNCKMGGLWRVDIPLKLVIM 629
```

```

QY 544 GIDCHDMTAGRSAGFVASINEGTRWFSCIFQDRGQELVDGLKYCLQALRAMNSC 603
DB 630 GIDCHDMTAGRSAGFVASINEGTRWFSCIFQDRGQELVDGLKYCLQALRAMNSC 689
QY 604 NEYMPESRIIVYRDGVDGOLKTLVNVYVPOFLDKLSIGRGYNPRLTVIVVKRVNTRFF 663
DB 690 NEYMPESRIIVYRDGVDGOLKTLVNVYVPOFLDKLSIGRGYNPRLTVIVVKRVNTRFF 749
QY 664 AQSGLRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSVSPSTHYNVYDNSGLKPDHIQ 723
DB 750 AQSGLRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSVSPSTHYNVYDNSGLKPDHIQ 809
QY 724 RLTYKLCCHIYNNWPGVIRVPAPCOYAHKLAFLVGQSIHREPNSLSNRLYYL 775
DB 810 RLTYKLCCHIYNNWPGVIRVPAPCOYAHKLAFLVGQSIHREPNSLSNRLYYL 861

RESULT 5
Q8TBYS PRELIMINARY; PRT; 861 AA.
AC Q8TBYS;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Piwi-like 1 (Drosophila).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028581; AAH28581.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003185; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 861 AA; 98545 MW; D33376EDED743A CRC64;

Query Match 60.5%; Score 469; DB 4; Length 861;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 769; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 GVNTQNLDHVKSCTGSSGIIIVRLSTNHFRLTSRPOWALYQVHIDYNPLMEARRLSAL 63
DB 90 GVNTQNLDHVKSCTGSSGIIIVRLSTNHFRLTSRPOWALYQVHIDYNPLMEARRLSAL 149
QY 64 LFQHEDLIGKCHAFDGTILFLPKELQKVTEVFSKTRNGEDVRITITLTNELPPTPTCL 123
DB 150 LFQHEDLIGKCHAFDGTILFLPKELQKVTEVFSKTRNGEDVRITITLTNELPPTPTCL 209
QY 124 QFNIIIPRLLKMNLOQIGRNYNPNPDIDISHRLVWPGFTTSILOYSIMLCTDV 183
DB 210 QFNIIIPRLLKMNLOQIGRNYNPNPDIDISHRLVWPGFTTSILOYSIMLCTDV 269
QY 184 SHKVLRSSETVLDFMNFYHOTEHKEHQVSKELIGLWLTCKYNNKTYRVDIDWDONPK 243
DB 270 SHKVLRSSETVLDFMNFYHOTEHKEHQVSKELIGLWLTCKYNNKTYRVDIDWDONPK 329
QY 244 STFKKADGSEVSFLYYRKQYNQIBITLQKPLVLSQPKRRRGPGGTLPGPAMLIPELCYL 303
DB 330 STFKKADGSEVSFLYYRKQYNQIBITLQKPLVLSQPKRRRGPGGTLPGPAMLIPELCYL 389
QY 304 TGLTDKMRNDFNMKDLAVHTRLTPEQRQREVGLIDYIHKQNVQRELWDGSLSPSNL 363
DB 390 TGLTDKMRNDFNMKDLAVHTRLTPEQRQREVGLIDYIHKQNVQRELWDGSLSPSNL 449
QY 364 LSPSGRILOTEKIHOGGKTFDYNPQFADWSKETRGAPLISVKNPLDNLIIYTRNRYEAA 423

```

```

DB 450 LSPSGRILOTEKIHOGGKTFDYNPQFADWSKETRGAPLISVKNPLDNLIIYTRNRYEAA 509
QY 424 SLIQNLKFKVTPAMQWQMKAIMIEVDRTETAYLVLQOKVTADTQIVVCLLSNRKDKYD 483
DB 510 SLIQNLKFKVTPAMQWQMKAIMIEVDRTETAYLVLQOKVTADTQIVVCLLSNRKDKYD 569
QY 484 AIKKYLCTDPTSPQCVVARTLGKQQTVMATATKIALQMNCKMGELWRVDIPLKLVMI 543
DB 570 AIKKYPTCDCTPSCQCVVARTLGKQQTVMATATKIALQMNCKMGELWRVDIPLKLVMI 629
QY 544 GIDCHDMTAGRSIAGFVASINEGTRWFSCIFQDRGQELVDGLKYCLQALRAMNSC 603
DB 630 GIDCHDMTAGRSIAGFVASINEGTRWFSCIFQDRGQELVDGLKYCLQALRAMNSC 689
QY 604 NEYMPESRIIVYRDGVDGOLKTLVNVYVPOFLDKLSIGRGYNPRLTVIVVKRVNTRFF 663
DB 690 NEYMPESRIIVYRDGVDGOLKTLVNVYVPOFLDKLSIGRGYNPRLTVIVVKRVNTRFF 749
QY 664 AQSGLRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSVSPSTHYNVYDNSGLKPDHIQ 723
DB 750 AQSGLRLQNLPGTVIDVEVTRPEWYDFIVSQAVRSVSPSTHYNVYDNSGLKPDHIQ 809
QY 724 RLTYKLCCHIYNNWPGVIRVPAPCOYAHKLAFLVGQSIHREPNSLSNRLYYL 775
DB 810 RLTYKLCCHIYNNWPGVIRVPAPCOYAHKLAFLVGQSIHREPNSLSNRLYYL 861

RESULT 6
Q9JMB7 PRELIMINARY; PRT; 862 AA.
AC Q9JMB7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE MIWI (piwi).
GN PIWI1 OR MIWI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_
RP SEQUENCE FROM N.A.
RA Miyagawa S.K., Kimura T., Nakano T.;
RT "Molecular Cloning and characterization of piwi family genes.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]_
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Deng W., Lin H.;
RT "Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
essential for spermatogenesis.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB032604; BAA93705.1; -.
DR EMBL; AF438405; AAL31014.1; -.
DR MGD; MGI:1928897; Piwi1.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003185; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 862 AA; 98574 MW; 45588D13284CCC4C CRC64;

Query Match 15.9%; Score 123; DB 11; Length 862;
Best Local Similarity 100.0%; Pred. No. 9.3e-120;
Matches 123; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 VLTYYNKKYRVDIDWDONPKSTFKKADGSEVSFLYYRKQYNQIBITLQKPLVLSQPK 281
DB 309 VLTYYNKKYRVDIDWDONPKSTFKKADGSEVSFLYYRKQYNQIBITLQKPLVLSQPK 368
QY 282 RRRPGGTLPGPAMLIPELCYLTGLTDKMRNDFNMKDLAVHTRLTPEQRQREVGLIDY 341

```



```
Db 369 RRRGGGTLPGFAMLIPELCYLTLGLTKQNRNDENVKDLAVHTRLPQQRQREVGLIDY 428
Qy 342 IHK 344
Db 429 IHK 431

RESULT 7
Q9GPA7 ID Q8UVX0 PRELIMINARY; PRT; 858 AA.
AC Q8UVX0;
DT 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Piwi protein.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Weeratne S.D., Gong Z., Tan C.-H.;
RT "Cloning and characterization of zebrafish homolog of piwi, essential
for germ-line stem cell self-renewal.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF336369; AAL57170.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 858 AA; 97451 MW; 6A12F2B511465777 CRC64;

Query Match 2.7%; Score 21; DB 13; Length 858;
Best Local Similarity 100.0%; Pred. No. 1.6e-12;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 741 RVPAPCQYAHKLAFLVQSGIH 761
Db 824 RVPAPCQYAHKLAFLVQSGIH 844

RESULT 8
Q9GPA7 ID Q9GPA7 PRELIMINARY; PRT; 808 AA.
AC Q9GPA7;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Seawi (Fragment).
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RA Rodriguez A.J., Bonder E.M.;
RT "Seawi - Cloning and Molecular Characterization of a Sea Urchin
Homolog of Piwi.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY014900; AAG42534.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 808 AA; 91366 MW; 19528F4B9D10474C CRC64;

Query Match 2.1%; Score 16; DB 5; Length 808;
Best Local Similarity 100.0%; Pred. No. 2.7e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 222 VLTKNKTYRVDDID 237
Db 302 VLTKNKTYRVDDID 317

RESULT 9
Q9GPA8 ID Q9GPA8 PRELIMINARY; PRT; 854 AA.
AC Q9GPA8;
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Seawi.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RA Rodriguez A.J., Bonder E.M.;
RT "Seawi - Cloning and Molecular Characterization of a Sea Urchin
Homolog of Piwi.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY014899; AAG42533.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 854 AA; 96722 MW; 509A1D39C0D1922C CRC64;

Query Match 2.1%; Score 16; DB 5; Length 854;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 222 VLTKNKTYRVDDID 237
Db 302 VLTKNKTYRVDDID 317

RESULT 10
Q17567 ID Q17567 PRELIMINARY; PRT; 722 AA.
AC Q17567;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE C01G5.2 protein.
GN C01G5.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA Bradshaw H., Stelliys L.;
RT "The sequence of C. elegans cosmid C01G5.";
RT Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
```

```

RN SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; U50068; AAB37734.1; -.
DR WormPep; COIGS.2; CE06748.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; PIWI.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; PIWI; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 722 AA; 82651 MW; 35BA0117952AD61F CRC64;

Query Match 1.9%; Score 15; DB 5; Length 722;
Best Local Similarity 100.0%; Pred. No. 2.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 RVPAPCOYAHKLAFL 755
DB 688 RVPAPCOYAHKLAFL 702

RESULT 11
P90786 AC P90786 PRELIMINARY; PRT; 824 AA.
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE D2030.6 protein.
GN D2030.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilkinson J.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851915;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z73906; CAA98113.1; -.
DR WormPep; D2030.6; CE09083.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; PIWI.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; PIWI; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 824 AA; 93844 MW; 970263CF6FDBCEB CRC64;

Query Match 1.9%; Score 15; DB 5; Length 824;
Best Local Similarity 100.0%; Pred. No. 3.2e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 RVPAPCOYAHKLAFL 755
DB 790 RVPAPCOYAHKLAFL 804

RESULT 12
Q8CC75 AC Q8CC75 PRELIMINARY; PRT; 121 AA.
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Epididymis;
RX MEDLINE=22354683; PubMed=1246851;
RA The FANTOM Consortium,
RT the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK033746; BAC28462.1; -.
KW Hypothetical protein.
SQ SEQUENCE 121 AA; 13704 MW; 9084DE582EBB8584 CRC64;

Query Match 1.4%; Score 11; DB 11; Length 121;
Best Local Similarity 100.0%; Pred. No. 0.0087;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 703 VSPTHYNVIYD 713
DB 61 VSPTHYNVIYD 71

RESULT 13
Q8N8G9 AC Q8N8G9 PRELIMINARY; PRT; 666 AA.
ID Q8N8G9
AC Q8N8G9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ39518 (piwi).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,
RA Wakebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RT "NED0 human cDNA sequencing project."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK096837; BAC04873.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; PIWI.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; PIWI; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
KW Hypothetical protein.
SQ SEQUENCE 666 AA; 76099 MW; 1082D86916390FA8 CRC64;

Query Match 1.4%; Score 11; DB 4; Length 666;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 742 VPAPOYAHKL 752
DB 633 VPAPOYAHKL 643

RESULT 14
Q8N9V8

```

```
DB 819 VPAPCOYAHKL 829
|||||
Search completed: December 10, 2003, 17:33:27
Job time : 46 secs

ID Q8N9V8 PRELIMINARY; PRT; 852 AA.
AC Q8N9V8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ36156 (piwi).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hotta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Takahashi-Fujii A., Oshina A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Negai K., Isogai T.;
RA Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RL "NEDO human cDNA sequencing project."
DR EMBL; AK093475; BAC04179.1; -
DR InterPro; IPR003100; PAZ
DR InterPro; IPR003185; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0821; PAZ; 1.
DR PROSITE; PS0822; PIWI; 1.
KW Hypothetical protein.
SQ SEQUENCE 852 AA; 96560 MW; 75C6FEFAE70701B CRC64;

Query Match 1.4%; Score 11; DB 4; Length 852;
Best Local Similarity 100.0%; Pred.No. 0.052;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 742 VPAPCOYAHKL 752
DB 819 VPAPCOYAHKL 829

RESULT 15
Q8NEH2 PRELIMINARY; PRT; 852 AA.
AC Q8NEH2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to piwi like homolog 1 (Drosophila).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC031060; AAH31060.1; -
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0821; PAZ; 1.
DR PROSITE; PS0822; PIWI; 1.
SQ SEQUENCE 852 AA; 96588 MW; 37769EE078B96D13 CRC64;

Query Match 1.4%; Score 11; DB 4; Length 852;
Best Local Similarity 100.0%; Pred.No. 0.052;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 742 VPAPCOYAHKL 752
```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 17:32:26 ; Search time 18 Seconds

(without alignments)
2024.761 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 775

Sequence: 1 MIFGVNTRQNLHDVKEKSTG.....VQGSIHREPNLSLNRLYL 775

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	1.2	1040	1 YO43 CAEEL	P34681 caenorhabdi
2	8	1.0	320	1 Y381 METJA	C45826 methanococ
3	8	1.0	369	1 WNT1 ANBME	P21551 ambystoma m
4	8	1.0	520	1 ECHE HELMO	O92968 heliobacill
5	7	0.9	112	1 YN48 ARCFC	O30321 archaeoglob
6	7	0.9	130	1 YJID ECOLI	P39375 escherichia
7	7	0.9	134	1 WNT1 CHICK	O91029 gallus gall
8	7	0.9	135	1 YJ01 AQUAE	O67739 aquifex aeo
9	7	0.9	148	1 VSN1 NEIMC	O9rlm5 neisseria m
10	7	0.9	149	1 VB15 VACCC	P21089 vaccinia vi
11	7	0.9	149	1 VB15 VACCV	P24772 vaccinia vi
12	7	0.9	149	1 VB15 VARV	P33877 variola vir
13	7	0.9	207	1 YFAT ECOLI	P76466 escherichia
14	7	0.9	209	1 ABPA PRUPE	O92ra4 prunus pers
15	7	0.9	209	1 ABGP PRUPE	O04012 prunus pers
16	7	0.9	209	1 ENG2 NEIGO	O05132 neisseria g
17	7	0.9	244	1 YBGL ECOLI	P75746 escherichia
18	7	0.9	274	1 DAPF HASIN	P44859 haemophilus
19	7	0.9	274	1 DAPF PASMU	P57962 pasteurella
20	7	0.9	274	1 DAPF SALTY	O916p6 salmonella
21	7	0.9	274	1 DAPF YERPE	P46357 versinia pe
22	7	0.9	275	1 CHER VIBAN	O57508 vibrio angu
23	7	0.9	275	1 CHER VIBCH	O5x9x2 vibrio para
24	7	0.9	275	1 CHRI VIBCH	O9kq06 vibrio chol
25	7	0.9	284	1 TLX2 HUMAN	O43763 homo sapien
26	7	0.9	284	1 TLX2 MOUSE	O61663 mus musculu
27	7	0.9	284	1 Y766 AQUAE	O66965 aquifex aeo
28	7	0.9	288	1 YQD5 CAEEL	O9285 caenorhabdi
29	7	0.9	291	1 TLX3 HUMAN	O43711 homo sapien
30	7	0.9	291	1 TLX3 MOUSE	O55144 mus musculu
31	7	0.9	294	1 SC17 PICPA	O9p4d0 pichia past
32	7	0.9	294	1 YJ05 YEAST	P47008 saccharomyc
33	7	0.9	297	1 TLX1 CHICK	O93366 gallus gall

34	7	0.9	297	1 TLX3 CHICK	O93367 gallus gall
35	7	0.9	319	1 PFSN ECOLI	P29131 escherichia
36	7	0.9	329	1 RRP4 SCHPO	O03704 schizosacch
37	7	0.9	330	1 TLX1 HUMAN	P31314 homo sapien
38	7	0.9	332	1 TLX1 MOUSE	P43345 mus musculu
39	7	0.9	335	1 AROB PYRFU	O8u0a8 pyrococcus
40	7	0.9	338	1 AROB PYRAB	O9v1h9 pyrococcus
41	7	0.9	356	1 VF7 BRD	P35935 broadhaven
42	7	0.9	357	1 CAD2 PICAB	O82035 picea abies
43	7	0.9	357	1 CAD7 PICAB	O83350 picea abies
44	7	0.9	357	1 CADH PINRA	O40976 pinus radia
45	7	0.9	357	1 CADH PINTA	P41637 pinus taeda

ALIGNMENTS

RESULT 1
YO43 CAEEL
ID YO43 CAEEL STANDARD; PRT; 1040 AA.
AC P34681;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK757.3 in chromosome III.
GN ZK757.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laiister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showken R.,
RA Sims M., Shalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z30215; CAA82941.1; -
CC EMBL; Z29121; CAA82941.1; JOINED.
CC EMBL; Z29121; CAA82389.1; -
CC EMBL; Z30215; CAA82389.1; JOINED.
CC PIR; D88568; D88568.
CC WormPep; ZK757.3; CE01117.
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003165; Piwi.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC PROSITE; P550821; PAZ; 1.
CC PROSITE; P550822; PIWI; 1.
KW Hypothetical protein.
FT DOMAIN 378 486 PAZ.

```

FT DOMAIN 660 966 PIWI.
SQ SEQUENCE 1040 AA; 115415 NW; 4C8483C1F1D72338 CRC64;

Query Match 1.2%; Score 9; DB 1; Length 1040;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 610 RIIVRDGV 618
DB 811 RIIVRDGV 819

RESULT 2
Y381_METJA STANDARD; PRT; 320 AA.
AC Q57826;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M70381.
GN M70381.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
SEQUENCE FROM N.A. / DSM 2661 / ATCC 43067;
RC STRAIN=JAL-1
RX MEDLINE=96337999; PubMed=8689087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Karvage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; U67491; AAB98370.1; -.
CC PIR; E64347; E64347.
CC TIGR; M70381; -.
CC InterPro; IPR002764; DUF73.
CC Pfam; PF01905; DUF73; 1.
CC ProDom; PD017873; DUF73; 1.
CC Hypothetical protein: Complete proteome.
SQ SEQUENCE 320 AA; 35924 MW; 81CAGDEI78F4B223 CRC64;

Query Match 1.0%; Score 8; DB 1; Length 320;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 KKADGSEV 254
DB 95 KKADGSEV 102

RESULT 3
WNT1_AMBME STANDARD; PRT; 369 AA.
AC P21551;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT

```

```

DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Wnt-1 protein precursor.
GN WNT-1.
OS Ambystoma mexicanum (Axolotl).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidae; Ambystomatidae;
OC Ambystoma.
OX NCBI_TaxID=8296;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=91081334; PubMed=2259633;
RA Busse U., Guay J., Seguin C.;
RT "Nucleotide sequence of a cDNA encoding Wnt-1 of the Mexican axolotl
RT Ambystoma mexicanum.";
RN Nucleic Acids Res. 18:7439-7439(1990).
RL [2]
ERRATUM.
RX MEDLINE=91204483; PubMed=2017393;
RA Busse U., Guay J., Seguin C.;
RL Nucleic Acids Res. 19:981-981(1991).
RN [3]
CHARACTERIZATION.
RX MEDLINE=93285407; PubMed=8508949;
RA Busse U., Seguin C.;
RT "Molecular analysis of the Wnt-1 proto-oncogene in Ambystoma
RT mexicanum (axolotl) embryos.";
RL Differentiation 53:7-15(1993).
CC -!- FUNCTION: LIGAND FOR MEMBERS OF THE FRIZZLED FAMILY OF SEVEN
CC TRANSMEMBRANE RECEPTORS. PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A
CC SIGNALING MOLECULE IMPORTANT IN CNS DEVELOPMENT. IS LIKELY TO
CC SIGNAL OVER ONLY A FEW CELL DIAMETERS.
CC -!- SUBCELLULAR LOCATION: Possibly secreted and associates with the
CC extracellular matrix.
CC -!- DEVELOPMENTAL STAGE: EARLY BLASTULA UNTIL GASTRULATION, BARELY
CC EXPRESSED DURING GASTRULATION AND PRESENT AGAIN FROM NEURULATION
CC UNTIL LATE EMBRYOGENESIS.
CC -!- SIMILARITY: Belongs to the Wnt family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; X55270; CAA38991.1; -.
CC PIR; S13721; S13721.
CC InterPro; IPR005817; Wnt.
CC InterPro; IPR005816; Wnt_grthfactor.
CC Pfam; PF00110; wnt; 1.
CC PRINTS; PR01349; WNTPROTEIN.
CC SMART; SM00097; WNT1; 1.
CC PROSITE; PS00246; WNT1; 1.
CC Wnt signaling pathway; Developmental protein; Glycoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 369
FT CARBOHYD 28 28
FT CARBOHYD 277 277
FT CARBOHYD 315 315
FT CARBOHYD 358 358
FT CARBOHYD 369 369
FT SEQUENCE 369 AA; 41383 MW; DC215A620F619321 CRC64;
SQ

Query Match 1.0%; Score 8; DB 1; Length 369;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 282 RRRPGCGT 289
DB 155 RRRPGCGT 162

RESULT 4

```

```

BCHB_HELMO
ID _BCHB_HELMO STANDARD; PRT; 520 AA.
AC Q9ZGE8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Light-independent protochlorophyllide reductase subunit B
DE (EC 1.18.-.-) (LI-FOR subunit B) (DPOR subunit B).
GN BCHB.

OS Helicobacillus mobilis.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
OC Helicobacillus.
OX NCBI_TaxID=28064;

RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=9061957; PubMed=9843979;
RA Xiong J., Inoue K., Bauer C.E.;
RT "Tracking molecular evolution of photosynthesis by characterization of
a major photosynthesis gene cluster from Helicobacillus mobilis.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:14851-14856(1998).
CC -!- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of
protochlorophyllide (pchlide) to form chlorophyllide a (Chlide)
(BY similarity). This reaction is light-independent.
CC -!- PATHWAY: Light-independent bacteriochlorophyll biosynthesis.
CC -!- SUBUNIT: Protochlorophyllide reductase is thought to be composed
of three subunits; bchL, bchN and bchB. Could form a
heterotrimer of two bchN and two bchN subunits.
CC -!- SIMILARITY: Belongs to the chlB / bchB / bchZ family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; AF080002; AAC84030.1; ALT_INIT.
FAMAP; MF_00353; -. 1.
InterPro; IPR000510; Oxred nitrognse1.
DR InterPro; IPR005969; Protochl reductB.
DR Pfam; PF00148; Oxidored nitro; 1.
DR TIGRFAMs; TIGR01278; DPOR BchB; 1.
OX Oxidoreductase, Photosynthesis; Bacteriochlorophyll biosynthesis.
SQ SEQUENCE 520 AA; 57775 MW; 7A2AC5F4AEC8A4BD CRC64;

Query Match 1.0%; Score 8; DB 1; Length 520;
Best Local Similarity 100.0%; Pred.No. 6.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 501 VARTLGKQ 508
DB 107 VARTLGKQ 114

RESULT 5
YN48_ARCFU STANDARD; PRT; 112 AA.
ID YN48_ARCFU
AC O30321;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AF2348.
GN AF2348.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
SEQUENCE FROM N.A.
RX STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RC MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

```

```

RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kervlavage A.R., Graham D.E., Kyripides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA Sadon P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; AB001114; AAB91315.1; -.
PIR; D69543; D69543.
DR TIGR; AF2348; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 112 AA; 13012 MW; C9C6AC0ACD6AC730 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 112;
Best Local Similarity 100.0%; Pred.No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 LIDYIHK 344
DB 34 LIDYIHK 40

RESULT 6
YJ1D_ECOLI STANDARD; PRT; 130 AA.
ID YJ1D_ECOLI
AC P39375;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yj1D.
GN YJ1D OR E4326.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
SEQUENCE FROM N.A.
RX STRAIN=KL12 / MG1655;
RC MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
region from 92.8 through 100 minutes";
RL Nucleic Acids Res. 23:2105-2119(1995).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; U14003; AAA97222.1; ALT_INIT.
DR EMBL; AE000503; AAC77282.1; ALT_INIT.
DR EcoGene; EG12565; YJ1D.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 130 AA; 14747 MW; AA07645C39525D90 CRC64;

```

Query Match 0.9%; Score 7; DB 1; Length 130;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 LRSALLF 65
| | | | |
DB 69 LRSALLF 75

RESULT 7

WNT1_CHICK STANDARD; PRT; 134 AA.
ID WNT1_CHICK AC Q91029;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Wnt-1 protein (Fragment).
GN WNT-1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White leghorn; TISSUE=Embryo;
RX MEDLINE=95121198; PubMed=7821210;
RA Bally-Cuif L., Wassef M.;
RT "Ectopic induction and reorganization of Wnt-1 expression in quail/chick chimeras".
RL Development 120:3379-3394 (1994).
CC -!- FUNCTION: LIGAND FOR MEMBERS OF THE FRIZZLED FAMILY OF SEVEN TRANSMEMBRANE RECEPTORS. PROBABLE DEVELOPMENTAL PROTEIN. MAY BE A SIGNALING MOLECULE IMPORTANT IN CNS DEVELOPMENT. IS LIKELY TO SIGNAL OVER ONLY FEW CELL DIAMETERS. PROMINENT ROLE IN THE INDUCTION OF THE MESENCEPHALON AND CEREBELLUM.
CC -!- SUBCELLULAR LOCATION: Possibly secreted and associates with the extracellular matrix.
CC -!- DEVELOPMENTAL STAGE: EXPRESSION IN THE MET-MESENCEPHALIC REGION.
CC -!- SIMILARITY: Belongs to the Wnt family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X81693; CAA57341.1; -;
PIR; I50729; I50729.
DR InterPro; IPR005817; Wnt.
DR InterPro; IPR005816; wnt_gthfactor.
DR Pfam; PF00110; wnt; 1.
DR PRINTS; PRO1349; WNTPROTEIN.
DR SMART; SM00097; WNT1; 1.
DR PROSITE; PS00246; WNT1; PARTIAL.
KW Wnt signaling pathway; Developmental protein; Glycoprotein.
FT NON_TER 1 1
FT TER 134 134
SQ SEQUENCE 134 AA; 14940 MW; 2E7D1E0DCFF01F8B CRC64;

Query Match 0.9%; Score 7; DB 1; Length 134;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 282 RRRGGG 288
| | | | |
DB 123 RRRGGG 129

RESULT 8

YJ01_AQUAE STANDARD; PRT; 135 AA.
ID YJ01_AQUAE AC O67739;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_1901.
GN AQ_1901.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L., Graham D.E., Overbeek R., Snead M.A., Kellar M., Aulay M., Huber R., Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus".
RL Nature 392:353-358 (1998).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE000762; AAC07709.1; -;
DR InterPro; IPR002716; PIN.
DR InterPro; IPR006596; PINC.
DR Pfam; PF01850; PIN; 1.
DR SMART; SM00670; PINC; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 135 AA; 15983 MW; 9C9B90271BFF267D CRC64;

Query Match 0.9%; Score 7; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 634 FLDCLS 640
| | | | |
DB 97 FLDCLS 103

RESULT 9

VSNI_NEIMC STANDARD; PRT; 148 AA.
ID VSNI_NEIMC AC Q9RLM5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative NmeDIP very-short-patch-repair endonuclease (EC 3.1.-.-)
DE (V.NmeDIP).
GN VSR OR NMEIDVP.
OS Neisseria meningitidis (serogroup C).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=135720;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2120 / Serogroup C / Serotype NT;
RX MEDLINE=20138154; PubMed=10671450;
RA Claus H., Friedrich A., Frosch M., Vogel U.;
RT "Differential distribution of novel restriction-modification systems in clonal lineages of Neisseria meningitidis".
RL J. Bacteriol. 182:1296-1303 (2000).
CC -!- FUNCTION: SPECULATED TO NICK NMEIDP SEQUENCES THAT CONTAIN T:G MISPAAIRS RESULTING FROM MSC-DEAMINATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE VSR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AJ238948; CAB59896.1; -;
 DR HSSP; P09184; 1VSR.
 DR REBASE; 4237; V.NmedIP.
 DR InterPro; IPR004603; Vsr.
 DR Pfam; PF03852; Vsr; 1.
 DR ProDom; PD016088; Vsr; 1.
 KW DNA repair; Hydrolase; Nuclease; Endonuclease; Restriction system.
 SQ SEQUENCE 148 AA; 17419 MW; D6AA99657B48499C CRC64;

Query Match 0.9%; Score 7; DB 1; Length 148;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 RLTPRO 331

Db 3 RLTPRO 9

RESULT 10

VB15 VACCV STANDARD; PRT; 149 AA.
 AC P21089; 1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein B15.
 GN B15R.
 OS Vaccinia virus (strain Copenhagen).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OC NCBI_TaxID=10249;

RP SEQUENCE FROM N.A.
 RX MEDLINE=91021027; PubMed=2219722;
 RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
 RA Paolletti E.;
 RL "The complete DNA sequence of vaccinia virus.";
 RL Virology 179:247-266(1990).
 RN COMPLETE GENOME.
 RP Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
 RA Paolletti E.;
 RT "Appendix to 'The complete DNA sequence of vaccinia virus'.";
 RL Virology 179:517-563(1990).
 CC -!- SIMILARITY: TO CAPRIPPOXVIRUS (STRAIN INS-1) AND SHOPE FIBROMA
 CC VIRUS PROTEINS T3A.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M35027; AAA48212.1; -;
 DR PIR; E42527; E42527.
 DR InterPro; IPR003867; Pox_B15.
 DR Pfam; PF02717; Pox_B15; 1.
 DE Early protein.
 GN B15R OR B14R OR B13R.
 OS Variola virus.
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

Query Match 0.9%; Score 7; DB 1; Length 149;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 YNIFRR 132
 Db 141 YNIFRR 147

RESULT 11

VB15 VACCV STANDARD; PRT; 149 AA.
 AC P24772;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-MAR-1992 (Rel. 21, Last sequence update)
 DT 01-MAY-1992 (Rel. 22, Last annotation update)
 DE Protein B15.
 GN B15R OR B14R.
 OS Vaccinia virus (strain WR).
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 OC NCBI_TaxID=10254;

RP SEQUENCE FROM N.A.
 RX MEDLINE=91259063; PubMed=2045793;
 RA Smith G.L., Chan Y.S., Howard S.T.;
 RT "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
 RT the right inverted terminal repeat.";
 RL J. Gen. Virol. 72:1349-1376(1991).
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91259063; PubMed=2045793;
 RA Smith G.L., Chan Y.S., Howard S.T.;
 RT "Nucleotide sequence of 42 kbp of vaccinia virus strain WR from near
 RT the right inverted terminal repeat.";
 RL J. Gen. Virol. 72:1349-1376(1991).
 RN [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=91111982; PubMed=1846491;
 RA Howard S.T., Chan Y.S., Smith G.L.;
 RT "Vaccinia virus homologues of the Shope fibroma virus inverted
 RT terminal repeat proteins and a discontinuous ORF related to the tumor
 RT necrosis factor receptor family.";
 RL Virology 180:633-647(1991).
 CC -!- SIMILARITY: TO CAPRIPPOXVIRUS (STRAIN INS-1) AND SHOPE FIBROMA
 CC VIRUS PROTEINS T3A.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; D11079; BAA01844.1; -;
 DR EMBL; M58053; AAA47964.1; -;
 DR PIR; JQ1808; JQ1808.
 DR InterPro; IPR003867; Pox_B15.
 DR Pfam; PF02717; Pox_B15; 1.
 DE Early protein.
 GN B15R OR B14R OR B13R.
 OS Variola virus.
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

SEQUENCE 149 AA; 17382 MW; A6118A960A0B6973 CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 149;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 YNIFRR 132
 Db 141 YNIFRR 147

RESULT 12

VB15 VARV STANDARD; PRT; 149 AA.
 AC P33877;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Protein B15.
 GN B15R OR B14R OR B13R.
 OS Variola virus.
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

Query Match 0.9%; Score 7; DB 1; Length 149;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OC Orthopoxvirus.
 OX NCBI_TaxID=10255;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=India-1967 / Isolate Ind3;
 RX MEDLINE=92209372; PubMed=1666548;
 RA Shchelkunov S.N., Marennikova S.S., Totmenin A.V., Blinov V.M.,
 RA Chizhikov V.E., Gutorov V.V., Saifonov P.F., Pozdnyakov S.G.,
 RA Shelukhina E.M., Gashnikov P.V., Anjaparidze O.G., Sandakchchiev L.S.;
 RT "Creation of a clone library of fragments from the natural variola
 virus and study of the structural and functional organization of
 RT viral genes from a circle of hosts";
 RL Dokl. Akad. Nauk SSSR 321:402-406(1991).
 RN [2]
 RP COMPLETE GENOME.
 RC STRAIN=India-1967 / Isolate Ind3;
 RX MEDLINE=93202281; PubMed=8384129;
 RA Shchelkunov S.N., Blinov V.M., Sandakchchiev L.S.;
 RT "Genes of variola and vaccinia viruses necessary to overcome the host
 RT protective mechanisms.";
 RL FEBS Lett. 319:80-83(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bangladesh-1975;
 RX MEDLINE=94088747; PubMed=8264798;
 RA Massung R.F., Esposito J.J., Liu L., Qi J., Utterback T.R.,
 RA Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.,
 RA Selivanov N.A., Cavallaro K.F., Kerlavage A.R., Mahy B.W.J.,
 RA Venter C.J.;
 RT "Potential virulence determinants in terminal regions of variola
 RT smallpox virus genome";
 RL Nature 366:748-751(1993).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Somalia-1977;
 RX MEDLINE=93202281; PubMed=8384129;
 RA Massung R.F., Loparev V.N., Knight J.C., Chizhikov S.N., Esposito J.J.,
 RA Totmenin A.V., Shchelkunov S.N., Parsons J.M.,
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X69198; CAA49123.1; -;
 DR EMBL; L22579; AAA60922.1; -;
 DR EMBL; X67117; CAA47524.1; -;
 DR EMBL; U18341; AAA69454.1; -;
 DR PIR; E36856; E36856.
 DR PIR; T28612; T28612.
 DR InterPro; IPR003867; Pox.B15.
 DR Pfam; PF02717; Pox.B15.1;
 SQ SEQUENCE 149 AA; 17369 MW; A84910243C7ED3D4 CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 149;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 126 YNIIIFRR 132
 Db 141 YNIIIFRR 147
 RESULT 13
 YFAT_ECOLI
 ID YFAT_ECOLI STANDARD; PRT; 207 AA.
 AC P76456;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein yfat precursor.
 GN YFAT OR B2229.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474(1997).
 CC -1- SIMILARITY: STRONG, TO P.AERUGINOSA PA4490 AND T.MARITIMA TM0986.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB000312; AAC75289.1; ALT_INIT.
 DR EcoGene; EGI4082; Yfat.
 KW Hypothetical protein; Signal; Complete proteome.
 FT SIGNAL 1 19 POTENTIAL,
 FT CHAIN 20 207 HYPOTHETICAL PROTEIN YFAT.
 SQ SEQUENCE 207 AA; 23767 MW; 6F3386370D78CECD CRC64;
 Query Match 0.9%; Score 7; DB 1; Length 207;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 326 LTPEQRQ 332
 Db 91 LTPEQRQ 97
 RESULT 14
 ABFA_PRUPE
 ID ABFA_PRUPE STANDARD; PRT; 209 AA.
 AC Q9ZRA4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Auxin-binding protein ABP19a precursor.
 OS ABP19A.
 OS Prunus persica (Peach).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
 OX NCBI_TaxID=3760;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Akatsuki; TISSUE=Shoot apex;
 RX MEDLINE=98329416; PubMed=9664714;
 RA Ohmura A., Tanaka Y., Kadowaki K., Hayashi T.;
 RT "Cloning of genes encoding auxin-binding proteins (ABP19/20) from
 RT peach: significant peptide sequence similarity with germin-like
 RT proteins.";
 RL Plant Cell Physiol. 39:492-499(1998).
 CC -1- FUNCTION: PROBABLE RECEPTOR FOR THE PLANT GROWTH-PROMOTING HORMONE
 CC AUXIN.
 CC -1- SUBUNIT: INTERACTS WITH ABP20.
 CC -1- SUBCELLULAR LOCATION: Cell wall.
 CC -1- SIMILARITY: BELONGS TO THE GERMIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U79114; AD00295.1; -.
DR HSP; P45850; IFI2.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin_sup.
DR InterPro; IPR001929; Germin.
DR Pfam; PF00190; Cupin; 1.
DR PRINTS; PR00325; GERMIN.
DR PROSITE; PS00725; GERMIN; 1.
DR Receptor; Apoplast; Cell wall; Signal; Glycoprotein; Manganese;
KW Metal-binding; Multigene family.
FT SIGNAL 1 18
FT CHAIN 19 209
FT METAL 101 101
FT METAL 103 103
FT METAL 108 108
FT METAL 147 147
FT METAL 147 147
FT DISULFID 24 39
FT CARBOHYD 60 60
SQ SEQUENCE 209 AA; 21859 MW; CAGE1F0A29A03414 CRC64;
POTENTIAL.
AUXIN-BINDING PROTEIN ABP19a.
MANGANESE (BY SIMILARITY).
MANGANESE (BY SIMILARITY).
MANGANESE (BY SIMILARITY).
MANGANESE (BY SIMILARITY).
BY SIMILARITY.
N-LINKED (GLCNAC. . .) (POTENTIAL).
SEQUENCE 209 AA; 21859 MW; CAGE1F0A29A03414 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 209;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 558 IAGFVAS 564
DB 118 IAGFVAS 124

RESULT 15
ID ABPB_PRUE STANDARD; PRT; 209 AA.
AC 004012;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Auxin-binding protein ABP19b precursor.
GN ABP19B.
OS Prunus persica (Peach).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
OX NCBI_TaxID=3760;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Akatsuki; TISSUE=Shoot apex;
RX MEDLINE=98329416; PubMed=9664714;
RA Ohmura A., Tanaka Y., Kadowaki K., Hayashi T.;
RT "Cloning of genes encoding auxin-binding proteins (ABP19/20) from
RT peach: significant peptide sequence similarity with germin-like
RT proteins.";
RL Plant Cell Physiol. 39:492-499(1998).
CC -!- FUNCTION: PROBABLE RECEPTOR FOR THE PLANT GROWTH-PROMOTING HORMONE
CC AUXIN.
CC -!- SUBUNIT: INTERACTS WITH ABP20.
CC -!- SUBCELLULAR LOCATION: Cell wall.
CC -!- SIMILARITY: BELONGS TO THE GERMIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; U01163; AAB51241.1; -.

DR HSP; P45850; IFI2.
DR InterPro; IPR006045; Cupin.
DR InterPro; IPR007113; Cupin_sup.
DR InterPro; IPR001929; Germin.
DR Pfam; PF00190; Cupin; 1.
DR PRINTS; PR00325; GERMIN.
DR PROSITE; PS00725; GERMIN; 1.
DR Receptor; Apoplast; Cell wall; Signal; Glycoprotein; Manganese;
KW Metal-binding; Multigene family.
FT SIGNAL 1 18
FT CHAIN 19 209
FT METAL 101 101
FT METAL 103 103
FT METAL 108 108
FT METAL 147 147
FT METAL 147 147
FT DISULFID 24 39
FT CARBOHYD 60 60
SQ SEQUENCE 209 AA; 21856 MW; 886FD412BD8DF771 CRC64;

Query Match 0.9%; Score 7; DB 1; Length 209;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 558 IAGFVAS 564
DB 118 IAGFVAS 124

Search completed: December 10, 2003, 17:33:55
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 17:32:31 ; Search time 21 Seconds
(without alignments)
3549.082 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 775
Sequence: 1 MIFGVNTRQNLHDHVKESKTG.....VGQSHREPNSLSNRLYYL 775

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 76:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	15	1.9	722	2 T30995	hypothetical prote
2	15	1.9	824	2 T20351	hypothetical prote
3	9	1.2	958	2 S41013	hypothetical prote
4	9	1.2	1032	2 T23164	hypothetical prote
5	9	1.2	1035	2 T23165	hypothetical prote
6	9	1.2	1040	2 D88568	protein ZK757.3 [i
7	8	1.0	320	2 E64347	hypothetical prote
8	8	1.0	369	2 S13721	Wnt-1 protein prec
9	8	1.0	398	2 T02484	hypothetical prote
10	8	1.0	528	2 T31459	protochlorophyllid
11	8	1.0	1137	2 T19414	hypothetical prote
12	7	0.9	61	2 AG1822	hypothetical prote
13	7	0.9	77	2 C71800	hypothetical prote
14	7	0.9	85	2 T44555	hypothetical prote
15	7	0.9	97	2 T31023	conserved hypothe
16	7	0.9	105	2 S27493	nodC protein - Bra
17	7	0.9	112	2 D69543	hypothetical prote
18	7	0.9	130	2 E87024	probable membrane
19	7	0.9	133	2 S65511	YjID protein - Esc
20	7	0.9	133	2 E91289	hypothetical prote
21	7	0.9	133	2 A86131	hypothetical prote
22	7	0.9	134	2 I50729	gene Wnt-1 protein
23	7	0.9	135	2 A10541	probable secreted
24	7	0.9	143	2 T37449	probable 16.7K pro
25	7	0.9	149	2 JQ1808	B14R 17.3K protein
26	7	0.9	149	2 E42527	B15R protein - vac
27	7	0.9	149	2 F72173	D3R protein - var
28	7	0.9	149	2 E36856	B14R protein - var
29	7	0.9	149	2 T28612	hypothetical prote

30 7 0.9 152 2 H95386 protein imported
31 7 0.9 164 2 T28741 hypothetical prote
32 7 0.9 165 2 T28742 hypothetical prote
33 7 0.9 188 2 B82696 cytochrome B561 XF
34 7 0.9 197 2 T28739 hypothetical prote
35 7 0.9 197 2 S74830 hypothetical prote
36 7 0.9 216 2 C64993 hypothetical prote
37 7 0.9 216 2 H91017 hypothetical prote
38 7 0.9 216 2 B85862 hypothetical prote
39 7 0.9 219 2 F71155 hypothetical prote
40 7 0.9 243 2 C96946 glutamine ABC tran
41 7 0.9 244 1 H64806 YsgL protein - Esc
42 7 0.9 244 2 B90721 probable lactam ut
43 7 0.9 244 2 H85571 probable lactam ut
44 7 0.9 251 2 A83660 hypothetical prote
45 7 0.9 266 2 A69358 carboxylesterase (

ALIGNMENTS

RESULT 1

T30995
hypothetical protein C01G5.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 29-Oct-1999
C:Accession: T30995
R:Bradshaw, H.; Stellyes, L.
submitted to the EMBL Data Library, August 1999
A:Description: The sequence of C. elegans cosmid C01G5.
A:Reference number: Z20956
A:Accession: T30995
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-722 <BRA>
A:Cross-references: EMBL:U50068; PIDN:AB37734.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Map position: IV
A:Introns: 58/3; 121/3; 202/3; 350/1; 471/3; 530/1; 609/2; 671/3
A:Note: C01G5.2

Query Match 1.9%; Score 15; DB 2; Length 722;
Best Local Similarity 100.0%; Pred. No. 1e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 RVPAPCOYAHKLAFL 755
Db 688 RVPAPCOYAHKLAFL 702

RESULT 2

T20351
hypothetical protein D2030.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20351
R:Wilkinson, J.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19261
A:Accession: T20351
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-824 <WTL>
A:Cross-references: EMBL:Z73906; PIDN:CAA98113.1; GSPDB:GN000019; CESP:D2030.6
A:Experimental source: clone D2030
C:Genetics:
A:Gene: CESP:D2030.6
A:Map position: 1
A:Introns: 118/3; 223/3; 304/3; 452/1; 573/3; 632/1; 711/2; 773/3
Query Match 1.9%; Score 15; DB 2; Length 824;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 RVPAPQYAKLAFL 755

Db 790 RVPAPQYAKLAFL 804

RESULT 3

S41013

hypothetical protein ZK757.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 18-Feb-2000

C:Accession: S41013

R:Thomas, K.

submitted to the EMBL Data Library, December 1993

A:Reference number: S41011

A:Accession: S41013

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-958 <THO>

A:Cross-references: EMBL:Z29121

C:Genetics:

A:Introns: 52/2; 350/2; 422/3; 592/1; 679/3; 721/2; 897/2; 929/1

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 1.2%; Score 9; DB 2; Length 958;

Best Local Similarity 100.0%; Pred. No. 2.5;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 610 RIIVYRDGV 618

Db 729 RIIVYRDGV 737

RESULT 4

T23164

hypothetical protein T22B3.2a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000

C:Accession: T23164; T25099

R:Cottage, A.

submitted to the EMBL Data Library, January 1996

A:Reference number: Z19701

A:Accession: T23164

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1032 <WIL>

A:Cross-references: EMBL:Z68750; PIDN:CAA92969.1; GSPDB:GN00022; CESP:T22B3.2a

A:Experimental source: clone K01A6

R:Lennard, N.

submitted to the EMBL Data Library, December 1995

A:Reference number: Z19981

A:Accession: T25099

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1032 <WIL>

A:Cross-references: EMBL:Z68300; PIDN:CAA92618.1; GSPDB:GN00022; CESP:T22B3.2a

A:Experimental source: clone T22B3

C:Genetics:

A:Gene: CESP:T22B3.2a

A:Map position: 4

A:Introns: 19/1; 71/3; 129/2; 424/2; 496/3; 607/1; 666/1; 753/3; 795/2; 961/2; 1003/1

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match

Best Local Similarity 1.2%; Score 9; DB 2; Length 1032;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 610 RIIVYRDGV 618

Db 803 RIIVYRDGV 811

RESULT 5

T23165

hypothetical protein T22B3.2b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000

C:Accession: T23165; T25100

R:Cottage, A.

submitted to the EMBL Data Library, January 1996

A:Reference number: Z19701

A:Accession: T23165

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1035 <WIL>

A:Cross-references: EMBL:Z68750; PIDN:CAA92970.1; GSPDB:GN00022; CESP:T22B3.2b

A:Experimental source: clone K01A6

R:Lennard, N.

submitted to the EMBL Data Library, December 1995

A:Reference number: Z19981

A:Accession: T25100

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1035 <WIL>

A:Cross-references: EMBL:Z68300; PIDN:CAA92619.1; GSPDB:GN00022; CESP:T22B3.2b

A:Experimental source: clone T22B3

C:Genetics:

A:Gene: CESP:T22B3.2b

A:Map position: 4

A:Introns: 19/1; 74/3; 132/2; 427/2; 499/3; 610/1; 669/1; 756/3; 798/2; 964/2; 1006/1

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match

Best Local Similarity 1.2%; Score 9; DB 2; Length 1035;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 610 RIIVYRDGV 618

Db 806 RIIVYRDGV 814

RESULT 6

D88568

protein ZK757.3 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001

C:Accession: D88568

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biol.

A:Reference number: A75000; MUID:99069613; PMID:9851916

A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/projects/C_e

A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; a

A:Accession: D88568

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1040 <STO>

A:Cross-references: GB:chr_III; PIDN:CAA82941.1; PID:93877004; GSPDB:GN00021; CESP:ZK7

C:Genetics:

A:Gene: ZK757.3

A:Map position: 3

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match

Best Local Similarity 1.2%; Score 9; DB 2; Length 1040;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 610 RIIVYRDGV 618

Db 811 RIIVYRDGV 819

RESULT 7

E64347

hypothetical protein MJ0381 - Methanococcus jannaschii

C;Species: Methanococcus jannaschii
 C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C;Accession: E64347
 R;Buit, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Reason, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. Science 273, 1058-1073, 1996
 A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C. A;Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii A;Reference number: A64300; MUID:96337999; PMID:8668087
 A;Accession: E64347
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: DNA
 A;Residues: 1-320 <BUL>
 A;Cross-references: GB:U67491; GB:L77117; NID:g2826278; PIDN:AAB98370.1; PID:g1591087; C;Genetics:
 A;Map position: FOR345977-346939
 C;Superfamily: Pyrococcus horikoshii hypothetical protein PH0920

Query Match 1.0%; Score 8; DB 2; Length 320;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 KKADGSEV 254
 |||||
 DB 95 KKADGSEV 102

RESULT 8
 S13721
 Wnt-1 protein precursor, secreted - axolotl
 C;Species: Ambystoma mexicanum (axolotl)
 C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
 C;Accession: S13721; S13722
 R;Busse, U.; Guay, J.; Seguin, C. Nucleic Acids Res. 18, 7439, 1990
 A;Title: Nucleotide sequence of a cDNA encoding Wnt-1 of the Mexican axolotl Ambystoma A;Reference number: S13721; MUID:91081334; PMID:2259633
 A;Accession: S13721
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown
 A;Molecule type: mRNA
 A;Residues: 1-369 <BUS>
 A;Cross-references: EMBL:X55270; NID:g62424; PIDN:CAA38991.1; PID:g62425
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1990
 R;Busse, U.; Guay, J.; Seguin, C. Nucleic Acids Res. 19, 981, 1991
 A;Title: Nucleotide sequence of a cDNA encoding Wnt-1 of the Mexican axolotl Ambystoma A;Reference number: S13722; MUID:91204483; PMID:2017393
 A;Contents: annotation; reprinted sequence figure
 C;Genetics:
 A;Gene: Wnt-1
 C;Superfamily: int-1 transforming protein

Query Match 1.0%; Score 8; DB 2; Length 369;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 282 RRRGPGGT 289
 |||||
 DB 155 RRRGPGGT 162

RESULT 9
 T02484
 hypothetical protein At2g30010 [imported] - Arabidopsis thaliana
 N;Alternate names: hypothetical protein F23F1.7
 C;Species: Arabidopsis thaliana (mouse-ear cress)
 C;Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
 C;Accession: T02484; D84703
 R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul submitted to the EMBL Data Library, August 1998
 A;Description: Arabidopsis thaliana chromosome II BAC F23F1 genomic sequence.
 A;Reference number: Z14675

A;Accession: T02484
 A;Status: translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-398 <ROU>
 A;Cross-references: EMBL:AC004680; NID:g3420043; PID:g3420050
 A;Experimental source: cultivar Columbia
 R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y. M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999
 A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A;Reference number: A84420; MUID:20083487; PMID:10617197
 A;Accession: D84703
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-398 <STO>
 A;Cross-references: GB:AE002093; NID:g3420050; PIDN:AAC31851.1; GSPDB:GN00139
 C;Genetics:
 A;Gene: At2G30010; F23F1.7
 A;Map position: 2
 A;Introns: 112/2; 169/3; 230/2; 283/2

Query Match 1.0%; Score 8; DB 2; Length 398;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 SVSPHYN 709
 |||||
 DB 276 SVSPHYN 283

RESULT 10
 T31459
 protochlorophyllide reductase (EC 1.3.1.33) chain B BchB, light-independent - Helioobac
 C;Species: Helioobacillus mobilis
 C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 15-Sep-2000
 C;Accession: T31459
 R;Xiong, J.; Inoue, K.; Bauer, C.E. Proc. Natl. Acad. Sci. U.S.A. 95, 14851-14856, 1998
 A;Title: Tracking molecular evolution of photosynthesis by characterization of a major A;Reference number: Z21036; MUID:99061957; PMID:9843979
 A;Accession: T31459
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-528 <XIO>
 A;Cross-references: EMBL:AF080002; NID:g3820536; PID:g3820557; PIDN:AAC84030.1
 C;Genetics:
 A;Gene: bchB
 C;Superfamily: protochlorophyllide reductase chain chlB
 C;Keywords: oxidoreductase

Query Match 1.0%; Score 8; DB 2; Length 528;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 501 VARTLGKQ 508
 |||||
 DB 115 VARTLGKQ 122

RESULT 11
 T19414
 hypothetical protein C23H4.6 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C;Accession: T19414
 R;Wilkinson, J. submitted to the EMBL Data Library, August 1996
 A;Reference number: Z19121
 A;Accession: T19414
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-1137 <WIL>

A;Cross-references: EMBL:Z78416; PIDN:CA801681.1; GSPDB:GN00028; CBSP:C23H4.6

A;Experimental source: clone C23H4

C;Genetics:

A;Gene: CBSP:C23H4.6

A;Map position: X

A;Introns: 34/1; 76/1; 111/2; 482/2; 598/2; 1091/3

Query Match 1.0%; Score 8; DB 2; Length 1137;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 ITITLTNE 114

DB 115 ITITLTNE 122

RESULT 12

AG1862 hypothetical protein asl0448 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AG1862

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AG1862

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-61 <KUR>

A;Cross-references: GB:BA000019; PIDN:BAB72406.1; PID:gl129793; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: asl0448

Query Match 0.9%; Score 7; DB 2; Length 61;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 512 MAIATKI 518

DB 36 MAIATKI 42

RESULT 13

C71800 hypothetical protein jhp1495 - Helicobacter pylori (strain J99)

C;Species: Helicobacter pylori

A;Variety: strain J99

C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C;Accession: C71800

R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;

Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;

Nature 397, 178-180, 1999

A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path

A;Reference number: A71800; MUID:99120557; PMID:9923682

A;Accession: C71800

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-77 <ARN>

A;Cross-references: GB:AE001571; GB:AE001439; NID:g4156120; PIDN:AAD07071.1; PID:g415612

A;Experimental source: strain J99

C;Genetics:

A;Gene: jhp1495

Query Match 0.9%; Score 7; DB 2; Length 77;

Best Local Similarity 100.0%; Pred. No. 32;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

217 LIGLVVL 223

DB 63 LIGLVVL 69

RESULT 14

T44555

hypothetical protein PA0631 [imported] - Pseudomonas aeruginosa (strain PA01)

C;Species: Pseudomonas aeruginosa

C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 06-Oct-2000

C;Accession: T44555; E83568

R;Nakayama, K.; Takashima, K.; Ishihara, H.; Shinomiya, T.; Kageyama, M.; Kanaya, S.;

submitted to the EMBL Data Library, August 1999

A;Description: Genetic relationship between bacteriophages and bacteriophages.

A;Reference number: Z22790

A;Accession: T44555

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-85 <NAK>

A;Cross-references: EMBL:AB030825; PIDN:BA083170.1

A;Experimental source: strain PA01

R;Stover, C.K.; Phan, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.F.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Li

; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pat

A;Reference number: AB2950; MUID:20437337; PMID:10984043

A;Accession: E83568

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-85 <STO>

A;Cross-references: GB:AE004498; GB:AE004091; NID:g9946491; PIDN:AAG04020.1; GSPDB:GN

A;Experimental source: strain PA01

C;Genetics:

A;Gene: PA0631

Query Match 0.9%; Score 7; DB 2; Length 85;

Best Local Similarity 100.0%; Pred. No. 35;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 593 LQAALRA 599

DB 77 LQAALRA 83

RESULT 15

T31023

conserved hypothetical protein 108 - Sulfolobus sp. plasmid pNOB8

C;Species: Sulfolobus sp.

C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C;Accession: T31023

R;She, Q.; Phan, H.; Garrett, R.A.; Albers, S.V.; Stedman, K.M.; Zillig, W.

Extremophiles 2, 417-425, 1998

A;Title: Genetic profile of pNOB8 from Sulfolobus: the first conjugative plasmid from

A;Reference number: Z20959; MUID:99044580; PMID:9827331

A;Accession: T31023

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-97 <SHE>

A;Cross-references: EMBL:AJ010405; NID:el351926; PID:el351945; PIDN:CAA09129.1

A;Experimental source: strain NOB8H2

C;Genetics:

A;Genome: plasmid pNOB8

Query Match 0.9%; Score 7; DB 2; Length 97;

Best Local Similarity 100.0%; Pred. No. 39;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 679 IDVEVTR 685

DB 60 IDVEVTR 66

Search completed: December 10, 2003, 17:35:29

Job time : 24 secs

Wed Dec 17:53:48 2003

us-10-043-77 2.oligo.rpr

Page 5

PT New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer

XX Claim 2; SEQ ID NO 2603; 265pp + Sequence Listing; English.

XX The invention relates to a human colon tumour expressed polynucleotide
 CC (i) encoding a polypeptide (ii), ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (ii) and (iii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 523 AA;

Query Match 67.4%; Score 522; DB 23; Length 523;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 522; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 VSLEYRKYNOEITDLKQPVLSQPKRRGGTLPGPAMLIPELCYLTGLTDKWRND 313
 DB 2 VSLEYRKYNOEITDLKQPVLSQPKRRGGTLPGPAMLIPELCYLTGLTDKWRND 61

QY 314 FNMKDLAVHTRLTPEQORQEVGRGLIDYTHKNDNVQRELDRWGLSPDNLSPSGRILQT 373
 DB 62 FNMKDLAVHTRLTPEQORQEVGRGLIDYTHKNDNVQRELDRWGLSPDNLSPSGRILQT 121

QY 374 EKHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVEAANSLIQNLFKVT 433
 DB 122 EKHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVEAANSLIQNLFKVT 181

QY 434 PANGMQRKAIMIEVDRTAYLRVLQOKVTADTQIVVCLLSNRKDKYDAIKKYLCTDC 493
 DB 182 PANGMQRKAIMIEVDRTAYLRVLQOKVTADTQIVVCLLSNRKDKYDAIKKYLCTDC 241

QY 494 PTPSQCVVARTLGKQQTVAIAIKIALQNMCKMGELWRVDIPLKLVMIIVGIDCYHDMTA 553
 DB 242 PTPSQCVVARTLGKQQTVAIAIKIALQNMCKMGELWRVDIPLKLVMIIVGIDCYHDMTA 301

QY 554 GRSIAGFVASINEGMRWFSRCIFQDRGOELVDGLKVCLOAALRAWNSCNEYMPRSRIIV 613
 DB 302 GRSIAGFVASINEGMRWFSRCIFQDRGOELVDGLKVCLOAALRAWNSCNEYMPRSRIIV 361

QY 614 YRDGVGGQKTLVNYEVPOFLDCLKSIGRGNPRLTVIVVKRVTNTRFPAQSGRLQNP 673
 DB 362 YRDGVGGQKTLVNYEVPOFLDCLKSIGRGNPRLTVIVVKRVTNTRFPAQSGRLQNP 421

QY 674 LPGVIVDEVTREPWYDFEIVSQAVRSGSVSPHYNVINDNSGLKPDHIOBLTVKLCIIY 733
 DB 422 LPGVIVDEVTREPWYDFEIVSQAVRSGSVSPHYNVINDNSGLKPDHIOBLTVKLCIIY 481

QY 734 YNPFQVIRVPAPCOYAHKLAFLVQGSIHREPNLSNRLYYL 775
 DB 482 YNPFQVIRVPAPCOYAHKLAFLVQGSIHREPNLSNRLYYL 523

RESULT 2

AA90235

ID AA90235 standard; Protein; 861 AA.

XX

AC AA90235;

XX

DT 29-AUG-2000 (first entry)

XX

DE Human piwi protein, designated hiwi.

XX

KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
 KW anaemia; immunodeficiency; male infertility; human.

XX Homo sapiens.

XX Location/Qualifiers
 PH Key
 FT Misc-difference 76
 FT /label= Leu, Ile
 FT /note= "encoded by NTA"
 FT Misc-difference 303
 FT /label= Leu, Ile
 FT /note= "encoded by NTA"
 FT Misc-difference 735
 FT /label= Leu, Ile
 FT /note= "encoded by NTA"

XX WO200032039-A1.

XX 08-JUN-2000.

XX 03-DEC-1999; 99WO-US28764.

XX 04-DEC-1998; 98US-0110901.

XX (UYDU-) UNIV DUKE.

XX Lin H;

XX WPI; 2000-412085/35.

XX N-PSDB; AAA07588.

XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
 PT therapy of diseases such as cancer and in various research and
 PT diagnostic applications

XX Claim 4; Page 189-194; 201pp; English.

XX This sequence represents the human piwi family protein, designated
 CC hiwi. The piwi family nucleic acids and polypeptides are used in gene
 CC therapy of diseases such as cancer and also in various research and
 CC diagnostic applications. The sequences can also be used to treat
 CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.

XX Sequence 861 AA;

Query Match 54.3%; Score 421; DB 21; Length 861;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 521; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 254 VSLEYRKYNOEITDLKQPVLSQPKRRGGTLPGPAMLIPELCYLTGLTDKWRND 313

DB 340 VSLEYRKYNOEITDLKQPVLSQPKRRGGTLPGPAMLIPELCYLTGLTDKWRND 399

QY 314 FNMKDLAVHTRLTPEQORQEVGRGLIDYTHKNDNVQRELDRWGLSPDNLSPSGRILQT 373

DB 400 FNMKDLAVHTRLTPEQORQEVGRGLIDYTHKNDNVQRELDRWGLSPDNLSPSGRILQT 459

QY 374 EKHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVEAANSLIQNLFKVT 433

DB 460 EKHQGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVEAANSLIQNLFKVT 519

QY 434 PANGMQRKAIMIEVDRTAYLRVLQOKVTADTQIVVCLLSNRKDKYDAIKKYLCTDC 493

DB 520 PANGMQRKAIMIEVDRTAYLRVLQOKVTADTQIVVCLLSNRKDKYDAIKKYLCTDC 579

QY 494 PTPSQCVVARTLGKQQTVAIAIKIALQNMCKMGELWRVDIPLKLVMIIVGIDCYHDMTA 553

DB 580 PTPSQCVVARTLGKQQTVAIAIKIALQNMCKMGELWRVDIPLKLVMIIVGIDCYHDMTA 639

QY 554 GRSIAGFVASINEGMRWFSRCIFQDRGOELVDGLKVCLOAALRAWNSCNEYMPRSRIIV 613

DB 640 GRSIAGFVASINEGMRWFSRCIFQDRGOELVDGLKVCLOAALRAWNSCNEYMPRSRIIV 699

SQ Sequence 130 AA;

Query Match 8.6%; Score 67; DB 21; Length 130;
 Best Local Similarity 100.0%; Pred. No. 8.4e-61; Length 130;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 549 HDMTAGRSIAGFVASINEGMTWFSRCIFQDGGELVDGLKVCLOALRAWNSCNEYMP 608
 D5 62 HDMTAGRSIAGFVASINEGMTWFSRCIFQDGGELVDGLKVCLOALRAWNSCNEYMP 121

QY 609 SRIVYR 615
 D5 122 SRIVYR 128

RESULT 5

ABP43865
 ID ABP43865 standard; Protein; 498 AA.

AC ABP43865;
 DT 26-FEB-2003 (first entry)
 DE Human mRNA sequence.

XX Neuroprotective; immunomodulator; cancer;
 KW cytotactic; anti-inflammatory; gene therapy; nutritional supplement;
 KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
 KW vulnary.

OS Homo sapiens.

XX WO200231111-A2.

XX 18-APR-2002.

XX 11-OCT-2001; 2001WO-US27760.

XX 12-OCT-2000; 2000US-0687527.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI; 2002-426278/45.

DR N-PSDB; ABQ61109.

XX New polypeptides and their encoded proteins, useful as nutritional
 PT sources or supplements, or in gene therapy, particularly for treating
 PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
 PT inflammation.

PS Claim 20; SEQ ID # 768; 357pp + sequence listing; English.

XX The invention relates to 446 newly isolated polynucleotide sequences.
 CC The activity of polynucleotides of the invention may be described as,
 CC vulnary, neuroprotective, immunomodulator, cytostatic and
 CC anti-inflammatory. Compositions comprising nucleic acids of the invention
 CC are useful for treating a mammalian subject, or as nutritional sources or
 CC supplements. These are useful in gene therapy, particularly for treating
 CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
 CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
 CC inflammation. The nucleic acids and polypeptides are also useful in
 CC diagnostic and research methods. The sequences given in records
 CC ABP43844-ABP43989 represent polypeptides encoded by polynucleotides of
 CC the invention.

CC NOTE: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 498 AA;

Query Match 1.4%; Score 11; DB 23; Length 498;
 Best Local Similarity 100.0%; Pred. No. 0.042;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 742 VPAPCOYAHKL 752

D5 465 VPAPCOYAHKL 475

RESULT 6

ABB62102
 ID ABB62102 standard; Protein; 866 AA.

XX ABB62102;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 13098.

XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEXE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL06205.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.

PS Disclosure; SEQ ID NO 13098; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 866 AA;

Query Match 1.2%; Score 9; DB 22; Length 866;

Best Local Similarity 100.0%; Pred. No. 8.4;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 615 RDGVGDGQL 623

D5 706 RDGVGDGQL 714

RESULT 7

ABG40429

ID ABG40429 standard; Peptide; 71 AA.

XX AC ABG40429;
 XX DT 19-AUG-2002 (first entry)
 XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 30094.
 XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.
 XX OS Homo sapiens.
 XX WO200186003-A2.
 XX PD 15-NOV-2001.
 XX PF 30-JAN-2001; 2001WO-US00665.
 XX PR 04-FEB-2000; 2000US-180312P.
 XX PR 26-MAY-2000; 2000US-207456P.
 XX PR 30-JUN-2000; 2000US-0608408.
 XX PR 03-AUG-2000; 2000US-0632366.
 XX PR 21-SEP-2000; 2000US-234687P.
 XX PR 27-SEP-2000; 2000US-236359P.
 XX PR 04-OCT-2000; 2000GB-0024263.
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2002-114183/15.
 XX DT Spatially-addressable set of single exon nucleic acid probes, used to
 XX PT measure gene expression in human lung samples -
 XX PS Claim 27; SEQ ID No 30094; 634pp; English.
 XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary

CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
 CC and hyaline membrane disease. The present sequence is a peptide/protein
 CC encoded by a single exon probe of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

Sequence 71 AA;

Query Match 1.0%; Score 8; DB 23; Length 71;

Best Local Similarity 100.0%; Pred. No. 9.2;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 521 QMNCCKMG 528

DB 60 QMNCCKMG 67

RESULT 8

ABG12914

ID ABG12914 standard; Protein; 140 AA.

XX AC ABG12914;

XX DT 13-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #12905.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS77101.

XX New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity -

XX Claim 20; SEQ ID No 43273; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 140 AA;

Query Match 1.0%; Score 8; DB 22; Length 140;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 748 YAHKLAFL 755

Db 113 YAHKLAFL 120

RESULT 9

AG20040

ID AAG20040 standard; Protein; 254 AA.

XX AC AAG20040;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 22072.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388;
PR 02-AUG-1999; 99US-0146389;
PR 03-AUG-1999; 99US-0147038;
PR 04-AUG-1999; 99US-0147204;
PR 04-AUG-1999; 99US-0147302;
PR 05-AUG-1999; 99US-0147192;
PR 05-AUG-1999; 99US-0147260;
PR 06-AUG-1999; 99US-0147303;
PR 06-AUG-1999; 99US-0147416;
PR 09-AUG-1999; 99US-0147493;
PR 09-AUG-1999; 99US-0147935;
PR 10-AUG-1999; 99US-0148171;
PR 11-AUG-1999; 99US-0148319;
PR 12-AUG-1999; 99US-0148341;
PR 13-AUG-1999; 99US-0148565;
PR 13-AUG-1999; 99US-0148684;
PR 16-AUG-1999; 99US-0149368;
PR 17-AUG-1999; 99US-0149175;
PR 18-AUG-1999; 99US-0149426;
PR 20-AUG-1999; 99US-0149722;
PR 20-AUG-1999; 99US-0149723;
PR 20-AUG-1999; 99US-0149929;
PR 23-AUG-1999; 99US-0149902;
PR 23-AUG-1999; 99US-0149930;
PR 25-AUG-1999; 99US-0150566;
PR 26-AUG-1999; 99US-0150884;
PR 26-AUG-1999; 99US-0151065;
PR 27-AUG-1999; 99US-0151066;
PR 27-AUG-1999; 99US-0151080;
PR 27-AUG-1999; 99US-0151080;
PR 30-AUG-1999; 99US-0151303;
PR 31-AUG-1999; 99US-0151438;
PR 01-SEP-1999; 99US-0151930;
PR 07-SEP-1999; 99US-0152363;
PR 10-SEP-1999; 99US-0153070;
PR 13-SEP-1999; 99US-0153758;
PR 15-SEP-1999; 99US-0154018;
PR 16-SEP-1999; 99US-0154039;
PR 20-SEP-1999; 99US-0154779;
PR 22-SEP-1999; 99US-0155139;
PR 23-SEP-1999; 99US-0155486;
PR 24-SEP-1999; 99US-0155659;
PR 28-SEP-1999; 99US-0156458;
PR 29-SEP-1999; 99US-0156596;
PR 04-OCT-1999; 99US-0157117;
PR 05-OCT-1999; 99US-0157253;
PR 06-OCT-1999; 99US-0157865;
PR 07-OCT-1999; 99US-0158029;
PR 08-OCT-1999; 99US-0158232;
PR 12-OCT-1999; 99US-0158369;
PR 13-OCT-1999; 99US-0159293;
PR 13-OCT-1999; 99US-0159294;
PR 13-OCT-1999; 99US-0159295;
PR 14-OCT-1999; 99US-0159329;
PR 14-OCT-1999; 99US-0159330;
PR 14-OCT-1999; 99US-0159331;
PR 14-OCT-1999; 99US-0159637;
PR 14-OCT-1999; 99US-0159638;
PR 16-OCT-1999; 99US-0159584;
PR 21-OCT-1999; 99US-0160741;
PR 21-OCT-1999; 99US-0160767;
PR 21-OCT-1999; 99US-0160768;
PR 21-OCT-1999; 99US-0160770;
PR 21-OCT-1999; 99US-0160814;
PR 21-OCT-1999; 99US-0160815;
PR 22-OCT-1999; 99US-0160980;
PR 22-OCT-1999; 99US-0160981;
PR 22-OCT-1999; 99US-0160989;
PR 25-OCT-1999; 99US-0161404;
PR 25-OCT-1999; 99US-0161405;
PR 25-OCT-1999; 99US-0161406;
PR 26-OCT-1999; 99US-0161359;
PR 26-OCT-1999; 99US-0161360;
PR 26-OCT-1999; 99US-0161361;
PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920;
PR 28-OCT-1999; 99US-0161992;
PR 28-OCT-1999; 99US-0161993;
PR 29-OCT-1999; 99US-0162142;
Query Match 1.0%; Score 8; DB 21; Length 254;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 702 SVSPTHYN 709
Db 132 SVSPTHYN 139
|||||
RESULT 10
RAG20039
ID AAG20039 standard; Protein; 271 AA.
XX AC AAG20039;
XX AC AAG20039;
DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 22071.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX FN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180;
PR 09-MAR-1999; 99US-0123548;
PR 23-MAR-1999; 99US-0125788;
PR 25-MAR-1999; 99US-0126264;
PR 29-MAR-1999; 99US-0126785;
PR 01-APR-1999; 99US-0127462;
PR 06-APR-1999; 99US-0128234;
PR 08-APR-1999; 99US-0128714;
PR 16-APR-1999; 99US-0129845;
PR 19-APR-1999; 99US-0130077;
PR 21-APR-1999; 99US-0130449;
PR 23-APR-1999; 99US-0130510;
PR 23-APR-1999; 99US-0130891;
PR 28-APR-1999; 99US-0131449;
PR 30-APR-1999; 99US-0132048;
PR 30-APR-1999; 99US-0132407;
PR 04-MAY-1999; 99US-0132484;
PR 05-MAY-1999; 99US-0132485;
PR 06-MAY-1999; 99US-0132486;
PR 06-MAY-1999; 99US-0132487;
PR 07-MAY-1999; 99US-0132863;
PR 11-MAY-1999; 99US-0134256;
PR 14-MAY-1999; 99US-0134218;
PR 14-MAY-1999; 99US-0134219;
PR 14-MAY-1999; 99US-0134221;
PR 14-MAY-1999; 99US-0134370;
PR 18-MAY-1999; 99US-0134768;
PR 19-MAY-1999; 99US-0134941;
PR 20-MAY-1999; 99US-0135124;
PR 21-MAY-1999; 99US-0135353;
PR 24-MAY-1999; 99US-0135629;
PR 25-MAY-1999; 99US-0136021;
PR 27-MAY-1999; 99US-0136392;
PR 28-MAY-1999; 99US-0136782;
PR 01-JUN-1999; 99US-0137222;
PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0143277.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147004.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

Best Local Similarity 1.0%; Score 8; DB 21; Length 271;
100.0%; Pred. No. 31;

```
Matches      8; Conservative      0; Mismatches      0; Indels      0; Gaps      0;
Qy          702 SVSPTHYN 709
Db          149 SVSPTHYN 156
|||||
RESULT 11
AAG20038
ID AAG20038 standard; Protein; 277 AA.
XX
AC AAG20038;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 22070.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX
XX 05-MAR-1999; 99US-0123180.
XX
XX 09-MAR-1999; 99US-0123548.
XX
XX 23-MAR-1999; 99US-0125788.
XX
XX 25-MAR-1999; 99US-0126264.
XX
XX 29-MAR-1999; 99US-0126785.
XX
XX 01-APR-1999; 99US-0127462.
XX
XX 06-APR-1999; 99US-0128234.
XX
XX 08-APR-1999; 99US-0128714.
XX
XX 16-APR-1999; 99US-0129845.
XX
XX 19-APR-1999; 99US-0130077.
XX
XX 21-APR-1999; 99US-0130449.
XX
XX 23-APR-1999; 99US-0130510.
XX
XX 28-APR-1999; 99US-0130891.
XX
XX 30-APR-1999; 99US-0131449.
XX
XX 30-APR-1999; 99US-0132048.
XX
XX 30-APR-1999; 99US-0132407.
XX
XX 04-MAY-1999; 99US-0132484.
XX
XX 05-MAY-1999; 99US-0132485.
XX
XX 06-MAY-1999; 99US-0132486.
XX
XX 07-MAY-1999; 99US-0132487.
XX
XX 07-MAY-1999; 99US-0132863.
XX
XX 11-MAY-1999; 99US-0132863.
XX
XX 14-MAY-1999; 99US-0132863.
XX
XX 14-MAY-1999; 99US-0134218.
XX
XX 14-MAY-1999; 99US-0134219.
XX
XX 14-MAY-1999; 99US-0134221.
XX
XX 14-MAY-1999; 99US-0134370.
XX
XX 18-MAY-1999; 99US-0134768.
XX
XX 19-MAY-1999; 99US-0134941.
XX
XX 20-MAY-1999; 99US-0135124.
XX
XX 21-MAY-1999; 99US-0135353.
XX
XX 24-MAY-1999; 99US-0135629.
XX
XX 25-MAY-1999; 99US-0136021.
XX
XX 27-MAY-1999; 99US-0136392.
XX
XX 28-MAY-1999; 99US-0136782.
XX
XX 01-JUN-1999; 99US-0137222.
XX
XX 03-JUN-1999; 99US-0137528.
XX
XX 04-JUN-1999; 99US-0137502.
XX
XX 07-JUN-1999; 99US-0137724.
XX
XX 08-JUN-1999; 99US-0138094.
XX
XX 10-JUN-1999; 99US-0138540.
XX
XX 10-JUN-1999; 99US-0138847.
XX
XX 14-JUN-1999; 99US-0139119.
XX
XX 16-JUN-1999; 99US-0139452.
XX
XX 16-JUN-1999; 99US-0139453.
XX
XX 17-JUN-1999; 99US-0139492.
XX
XX 18-JUN-1999; 99US-0139454.
XX
XX 18-JUN-1999; 99US-0139455.
XX
XX 18-JUN-1999; 99US-0139456.
XX
XX 18-JUN-1999; 99US-0139457.
XX
XX 18-JUN-1999; 99US-0139458.
XX
XX 18-JUN-1999; 99US-0139459.
XX
XX 18-JUN-1999; 99US-0139460.
XX
XX 18-JUN-1999; 99US-0139461.
XX
XX 18-JUN-1999; 99US-0139462.
XX
XX 18-JUN-1999; 99US-0139463.
XX
XX 18-JUN-1999; 99US-0139750.
XX
XX 21-JUN-1999; 99US-0139763.
XX
XX 22-JUN-1999; 99US-0139817.
XX
XX 23-JUN-1999; 99US-0139899.
XX
XX 23-JUN-1999; 99US-0140353.
XX
XX 24-JUN-1999; 99US-0140354.
XX
XX 28-JUN-1999; 99US-0140695.
XX
XX 28-JUN-1999; 99US-0140823.
XX
XX 29-JUN-1999; 99US-0140991.
XX
XX 30-JUN-1999; 99US-0141287.
XX
XX 01-JUL-1999; 99US-0141842.
XX
XX 02-JUL-1999; 99US-0142154.
XX
XX 06-JUL-1999; 99US-0142055.
XX
XX 06-JUL-1999; 99US-0142390.
XX
XX 08-JUL-1999; 99US-0142803.
XX
XX 09-JUL-1999; 99US-0142920.
XX
XX 12-JUL-1999; 99US-0142977.
XX
XX 13-JUL-1999; 99US-0143542.
XX
XX 14-JUL-1999; 99US-0143624.
XX
XX 15-JUL-1999; 99US-0144005.
XX
XX 16-JUL-1999; 99US-0144085.
XX
XX 16-JUL-1999; 99US-0144086.
XX
XX 19-JUL-1999; 99US-0144325.
XX
XX 19-JUL-1999; 99US-0144331.
XX
XX 19-JUL-1999; 99US-0144332.
XX
XX 19-JUL-1999; 99US-0144333.
XX
XX 19-JUL-1999; 99US-0144334.
XX
XX 19-JUL-1999; 99US-0144335.
XX
XX 20-JUL-1999; 99US-0144352.
XX
XX 20-JUL-1999; 99US-0144632.
XX
XX 20-JUL-1999; 99US-0144884.
XX
XX 21-JUL-1999; 99US-0144814.
XX
XX 21-JUL-1999; 99US-0145086.
XX
XX 21-JUL-1999; 99US-0145088.
XX
XX 22-JUL-1999; 99US-0145085.
XX
XX 22-JUL-1999; 99US-0145087.
XX
XX 22-JUL-1999; 99US-0145089.
XX
XX 22-JUL-1999; 99US-0145192.
XX
XX 23-JUL-1999; 99US-0145145.
XX
XX 23-JUL-1999; 99US-0145218.
XX
XX 23-JUL-1999; 99US-0145224.
XX
XX 26-JUL-1999; 99US-0145276.
XX
XX 27-JUL-1999; 99US-0145913.
XX
XX 27-JUL-1999; 99US-0145918.
XX
XX 27-JUL-1999; 99US-0145919.
XX
XX 28-JUL-1999; 99US-0145951.
XX
XX 02-AUG-1999; 99US-0146386.
XX
XX 02-AUG-1999; 99US-0146388.
XX
XX 02-AUG-1999; 99US-0146389.
XX
XX 03-AUG-1999; 99US-0147038.
XX
XX 04-AUG-1999; 99US-0147204.
XX
XX 04-AUG-1999; 99US-0147302.
XX
XX 05-AUG-1999; 99US-0147192.
XX
XX 05-AUG-1999; 99US-0147260.
XX
XX 06-AUG-1999; 99US-0147303.
XX
XX 06-AUG-1999; 99US-0147416.
XX
XX 09-AUG-1999; 99US-0147493.
XX
XX 09-AUG-1999; 99US-0147935.
XX
XX 10-AUG-1999; 99US-0148171.
XX
XX 11-AUG-1999; 99US-0148319.
XX
XX 12-AUG-1999; 99US-0148341.
```

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.0%; Score 8; DB 21; Length 277;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 SVSPHTYN 709
Db 155 SVSPHTYN 162

RESULT 12
AAG48143
ID AAG48143 standard; Protein; 277 AA.
XX AC AAG48143;
XX DI 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 60766.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142377.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.

PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.0%; Score 8; DB 21; Length 277;
Best Local Similarity 100.0%; Pred.No. 32;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 702 SVSPHYN 709
DB 155 SVSPHYN 162

RESULT 13
AAB92722
ID AAB92722 standard; Protein; 371 AA.
XX
AC AAB92722;
XX
DT 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:11144.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW Homo sapiens.
OS Arabidopsis thaliana protein fragment SEQ ID NO: 60765.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
PN EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300353.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesising polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX Claim 8; SEQ ID 11144; 2537pp + CD ROM; English.
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX Sequence 371 AA;
SQ
Query Match 1.0%; Score 8; DB 22; Length 371;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 748 YAHKLAFL 755
Db 344 YAHKLAFL 351
RESULT 14
AAG48142
ID AAG48142 standard; Protein; 398 AA.
XX
XX AAG48142;
XX

DT 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 60765.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 18-MAY-1999; 99US-0134370.
XX 19-MAY-1999; 99US-0134768.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137528.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139453.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.

```
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 29-JUN-1999; 99US-0141287.
PR 30-JUN-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 01-JUL-1999; 99US-0142055.
PR 02-JUL-1999; 99US-0142390.
PR 06-JUL-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142920.
PR 09-JUL-1999; 99US-0142977.
PR 12-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0143624.
PR 14-JUL-1999; 99US-0144005.
PR 15-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0145986.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149502.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.0%; Score 8; DB 21; Length 398;
Best Local Similarity 100.0%; Pred.No. 45;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 702 SVSPHYN 709
Db 276 SVSPHYN 283

RESULT 15
AAG48141
ID AAG48141 standard; Protein; 433 AA.
XX
AC AAG48141;
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 60764.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
```


XX Arabidopsis thaliana.
XX EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0128264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.

```
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.0%; Score 8; DB 21; Length 433;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 702 SVSPTHYN 709
Db 311 SVSPTHYN 318
```

Search completed: December 10, 2003, 17:34:55
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: December 4, 2003, 19:11:53 ; Search time 41 Seconds
(without alignments)
4877.824 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 4102

Sequence: 1 MIFGVNTRQNLHVKEKSTG.....VQSIHREPLNSLNRLYYL 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.23.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	4102	100.0	775	4	Q96JD5		Q96jd5 homo sapien
2	4087	99.6	861	4	Q96J94		Q96j94 homo sapien
3	4072	99.3	861	4	Q8TBY5		Q8tby5 homo sapien
4	4071	99.2	861	4	Q95404		Q95404 homo sapien
5	3980	97.0	862	11	Q9JMB7		Q9jmb7 mus musculus
6	3883	94.7	829	4	Q8NA60		Q8na60 homo sapien
7	2828.5	69.0	858	13	Q8UVX0		Q8uvx0 brachydanio
8	2103.5	51.3	852	4	Q8NEH2		Q8neh2 homo sapien
9	2101.5	51.2	852	4	Q8N9V8		Q8n9v8 homo sapien
10	2009.5	49.0	854	5	Q9GPA8		Q9gpa8 strongyloce
11	1859.5	45.3	878	11	Q8CGT6		Q8cgt6 mus musculus
12	1838.5	44.8	866	4	Q8N8G9		Q8n8g9 homo sapien
13	1811.5	44.2	808	5	Q9GPA7		Q9gpa7 strongyloce
14	1766	43.1	971	11	Q9JMB6		Q9jmb6 mus musculus
15	1764	43.0	971	11	Q8CDG1		Q8cdg1 mus musculus
16	1754	42.8	973	4	Q8TC59		Q8tc59 homo sapien

17	1468	35.8	866	5	Q76922		O76922 drosophila
18	1428.5	34.8	843	5	Q9VKM1		Q9vkm1 drosophila
19	1389.5	33.9	580	11	Q99MV6		Q99mv6 mus musculus
20	1295.5	31.6	824	5	P90786		P90786 caenorhabdi
21	1230	30.0	530	4	Q96SW6		Q96sw6 homo sapien
22	1209	29.5	421	5	Q9GPA6		Q9gpa6 strongyloce
23	1162	28.3	722	5	Q17567		Q17567 caenorhabdi
24	989	24.1	791	5	Q8ISG8		Q8isg8 stylonychia
25	962	23.5	371	4	Q9NW28		Q9nw28 homo sapien
26	804.5	19.6	780	5	Q8MXZ9		Q8mxz9 tetrahymena
27	803	19.6	779	5	Q8MQL1		Q8mql1 tetrahymena
28	728	17.7	781	5	Q9USC9		Q9usc9 paramecium
29	520.5	12.7	1194	10	Q9SHF2		Q9shf2 arabidopsis
30	508.5	12.4	860	11	Q8CJG0		Q8cjd0 mus musculus
31	508.5	12.4	1014	10	Q9SHF3		Q9shf3 arabidopsis
32	507	12.4	978	10	Q8LP00		Q8lp00 oryza sativ
33	505.5	12.3	910	5	O16720		O16720 caenorhabdi
34	504.5	12.3	861	11	Q8CJF8		Q8cjf8 mus musculus
35	502.5	12.3	951	11	Q8BTF4		Q8btf4 mus musculus
36	502	12.2	860	11	Q8CJF9		Q8cjf9 mus musculus
37	500	12.2	950	5	Q9V3L2		Q9v3l2 drosophila
38	500	12.2	984	5	Q9V6V6		Q9v6v6 drosophila
39	499.5	12.2	857	11	Q8CJG1		Q8cjd1 mus musculus
40	498.5	12.2	1002	5	Q20578		Q20578 caenorhabdi
41	497.5	12.1	924	10	Q9ZVD5		Q9zvd5 arabidopsis
42	497	12.1	860	4	Q8H9G7		Q8h9g7 homo sapien
43	495.5	12.1	924	4	Q9HCK5		Q9hck5 homo sapien
44	495	12.1	848	11	Q8CGU0		Q8cgu0 mus musculus
45	491.5	12.0	1123	10	Q9LP83		Q9lp83 arabidopsis

ALIGNMENTS

RESULT 1

Q96JD5	PRELIMINARY;	PRT;	775 AA.
AC Q96JD5;			
DT 01-DEC-2001 (Tremblrel. 19, Created)			
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)			
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)			
DE HIWI.			
OS Homo sapiens (Human)			
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=Testis;			
RA Sharma A.K., Nelson M.C., Brandt J.E., Wessman M., Muhmud N.,			
RA Weller K.P., Hoffman R.,			
RT "Human CD34+ Stem Cells Express the hiwi Gene, a Human Homolog of the			
RT Drosophila Gene piwi.";			
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.			
DR EMBL; AF264004; AAK92281.1; -			
DR InterPro; IPR003100; PAZ;			
DR InterPro; IPR003165; Pwi1.			
DR Pfam; PF02170; PAZ; 1.			
DR Pfam; PF02171; Pwi1; 1.			
DR PROSITE; PS50821; PAZ; 1.			
DR PROSITE; PS50822; Pwi1; 1.			
SQ SEQUENCE 775 AA; 89484 MW; DF169A2E9EAFD916 CRC64;			

Query Match	100.0%;	Score 4102;	DB 4;	Length 775;
Best Local Similarity	100.0%;	Pred. No. 4.1e-318;		
Matches 775;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1 MIFGVNTRQNLHVKEKSTGSGIIVRLSTNHFRTSPQWALYQYHIDYNPLMEARRLR 60				
Db 1 MIFGVNTRQNLHVKEKSTGSGIIVRLSTNHFRTSPQWALYQYHIDYNPLMEARRLR 60				
QY 61 SALLFOHEDLCKCHAFDGTILFLPKLQOKVTFVPSKTRNGEDVRIITLTNELPPTSP 120				

```
Db 61 SALLFOHEDLIGKCHAFDGTILFLPKRLOOKVTEVFSKTRNGEDVRITITITLNEIPPTSP 120
Qy 121 TCFQFYNIIFRLLKIMNLQOQGRNYNPNNDPIDIPSHRLVIWPGFTTTSILOYENSIMLC 180
Db 121 TCFQFYNIIFRLLKIMNLQOQGRNYNPNNDPIDIPSHRLVIWPGFTTTSILOYENSIMLC 180
Qy 181 TDVSHKVLRSSETVLDPMFNFYHQTEBHKFOEQVSKELIGLVLTLYKNNKTYRVDIDWDQ 240
Db 181 TDVSHKVLRSSETVLDPMFNFYHQTEBHKFOEQVSKELIGLVLTLYKNNKTYRVDIDWDQ 240
Qy 241 NPKSFVKADGSEVSLEYRKOYNOEITDLKOPVLVSOPKRRGPGGTLPGPAMLIPEL 300
Db 241 NPKSFVKADGSEVSLEYRKOYNOEITDLKOPVLVSOPKRRGPGGTLPGPAMLIPEL 300
Qy 301 CYLTGLTDRNDNFNMKDLAVHTRLTTPQOREVGRLLIDYIHKNDNVORELDRDGLSPD 360
Db 301 CYLTGLTDRNDNFNMKDLAVHTRLTTPQOREVGRLLIDYIHKNDNVORELDRDGLSPD 360
Qy 361 SNLLSFGRILOTEKIHOGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVE 420
Db 361 SNLLSFGRILOTEKIHOGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVE 420
Qy 421 AANSLIQNLFKVTPANGMQRKAIMLEVDDRTYALVLOKVTADTQIVWCLLSNRKD 480
Db 421 AANSLIQNLFKVTPANGMQRKAIMLEVDDRTYALVLOKVTADTQIVWCLLSNRKD 480
Qy 481 KYDAIKKYLCTDPTSCQCVARTLGKQQTVMATKIALQNNCKMGELWVDIPLKLV 540
Db 481 KYDAIKKYLCTDPTSCQCVARTLGKQQTVMATKIALQNNCKMGELWVDIPLKLV 540
Qy 541 MIVGIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAW 600
Db 541 MIVGIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAW 600
Qy 601 NSCNEPMPSRIIYRDGVGGQKTLVNYEVQFLDCLKSIGRGNPRLTVIWKRVNT 660
Db 601 NSCNEPMPSRIIYRDGVGGQKTLVNYEVQFLDCLKSIGRGNPRLTVIWKRVNT 660
Qy 661 RFFAQSGRLQNPFGTVIDVEVTRPEWYDFFIQSQAVRSVSPHYNVIYDNSGLKPD 720
Db 661 RFFAQSGRLQNPFGTVIDVEVTRPEWYDFFIQSQAVRSVSPHYNVIYDNSGLKPD 720
Qy 721 HIQRUTYKLCCHIYNNWPGVIRVPAPCOYAHKLAFLVGQSIHREPNSLSNRLYYL 775
Db 721 HIQRUTYKLCCHIYNNWPGVIRVPAPCOYAHKLAFLVGQSIHREPNSLSNRLYYL 775

RESULT 2
Q96J94 ID Q96J94 PRELIMINARY; PRT; 861 AA.
AC Q96J94;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE PIWI protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Sha J.H.;
RT "Cloning and identification of human piwi protein related to testis
development."
RL Submitted (May-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF387507; AAK69348.1; -.
DR InterPro; IPR003100; PAZ.
DR InPro; IPR003165; PIWI.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; PIWI; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
```

```
SQ SEQUENCE 861 AA; 98603 MW; 58D7F6C7321DEFA4 CRC64;
Query Match
Best Local Similarity 99.6%; Score 4087; DB 4; Length 861;
Matches 772; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GVNTRQNLDRHVKESKGTSSGIIIVRLSTNNHFLTRSPQWALYQYHIDYNPLMEARLSAL 63
Db 90 GVNTRQNLDRHVKESKGTSSGIIIVRLSTNNHFLTRSPQWALYQYHIDYNPLMEARLSAL 149
Qy 64 LFOHEDLIGKCHAFDGTILFLPKRLOOKVTEVFSKTRNGEDVRITITITLNEIPPTSPCL 123
Db 150 LFOHEDLIGKCHAFDGTILFLPKRLOOKVTEVFSKTRNGEDVRITITITLNEIPPTSPCL 209
Qy 124 QFYNIIFRLLKIMNLQOQGRNYNPNNDPIDIPSHRLVIWPGFTTTSILOYENSIMLCIDV 183
Db 210 QFYNIIFRLLKIMNLQOQGRNYNPNNDPIDIPSHRLVIWPGFTTTSILOYENSIMLCIDV 269
Qy 184 SHKVLRSSTVLDPMFNFYHQTEBHKFOEQVSKELIGLVLTLYKNNKTYRVDIDWDQNP 243
Db 270 SHKVLRSSTVLDPMFNFYHQTEBHKFOEQVSKELIGLVLTLYKNNKTYRVDIDWDQNP 329
Qy 244 STPKKADGSEVSLEYRKOYNOEITDLKOPVLVSOPKRRGPGGTLPGPAMLIPELCYL 303
Db 330 STPKKADGSEVSLEYRKOYNOEITDLKOPVLVSOPKRRGPGGTLPGPAMLIPELCYL 389
Qy 304 TGLTDKMDNFNMKDLAVHTRLTTPQOREVGRLLIDYIHKNDNVORELDRDGLSFDNL 363
Db 390 TGLTDKMDNFNMKDLAVHTRLTTPQOREVGRLLIDYIHKNDNVORELDRDGLSFDNL 449
Qy 364 LSPSGRILOTEKIHOGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVEAAN 423
Db 450 LSPSGRILOTEKIHOGGKTFDYNPQFADWSKETRGAPLISVKPLDNWLLIYTRNVEAAN 509
Qy 424 SLIQNLFKVTPANGMQRKAIMLEVDDRTYALVLOKVTADTQIVWCLLSNRKDYD 483
Db 510 SLIQNLFKVTPANGMQRKAIMLEVDDRTYALVLOKVTADTQIVWCLLSNRKDYD 569
Qy 484 AIKKYLCTDPTSCQCVARTLGKQQTVMATKIALQNNCKMGELWVDIPLKLVIV 543
Db 570 AIKKYLCTDPTSCQCVARTLGKQQTVMATKIALQNNCKMGELWVDIPLKLVIV 629
Qy 544 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAWNSC 603
Db 630 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAWNSC 689
Qy 604 NEYMPSPRIIYRDGVGGQKTLVNYEVQFLDCLKSIGRGNPRLTVIWKRVNTRFF 663
Db 690 NEYMPSPRIIYRDGVGGQKTLVNYEVQFLDCLKSIGRGNPRLTVIWKRVNTRFF 749
Qy 664 AQSGRLQNPFGTVIDVEVTRPEWYDFFIQSQAVRSVSPHYNVIYDNSGLKPDHIQ 723
Db 750 AQSGRLQNPFGTVIDVEVTRPEWYDFFIQSQAVRSVSPHYNVIYDNSGLKPDHIQ 809
Qy 724 RITYKLCCHIYNNWPGVIRVPAPCOYAHKLAFLVGQSIHREPNSLSNRLYYL 775
Db 810 RITYKLCCHIYNNWPGVIRVPAPCOYAHKLAFLVGQSIHREPNSLSNRLYYL 861

RESULT 3
Q8TBYS ID Q8TBYS PRELIMINARY; PRT; 861 AA.
AC Q8TBYS;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Piwi-like 1 (Drosophila).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RC TISSUE-Testis;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028581; AAH28581.1; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PSS0821; PAZ; 1.
DR PROSITE; PSS0822; PIWI; 1.
SQ SEQUENCE 861 AA; 98545 MW; D33376EDED743A CRC64;

Query Match 99.3%; Score 4072; DB 4; Length 861;
Best Local Similarity 99.6%; Pred. No. 8.4e-316;
Matches 769; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVNTRQNLHVKESKTGSSGIIIVRLSTNNHRLTSRQWALYQYHIDYNPMEARRLSAL 63
Db 90 GVNTRQNLHVKESKTGSSGIIIVRLSTNNHRLTSRQWALYQYHIDYNPMEARRLSAL 149

QY 64 LFOHEDLIGKCHAFDGTILFLPKRLQKQVTEVFSKTRNGEDVRITITLTNELPPTSPCTL 123
Db 150 LFOHEDLIGKCHAFDGTILFLPKRLQKQVTEVFSKTRNGEDVRITITLTNELPPTSPCTL 209

QY 124 QFYNIIFRLLKIMNLQOIGRNYNPNPIDIPSHRLVWPFGTTSIIQYENSIMLCTDV 183
Db 210 QFYNIIFRLLKIMNLQOIGRNYNPNPIDIPSHRLVWPFGTTSIIQYENSIMLCTDV 269

QY 184 SHKVLRSVLDPMFNFYHQTTEHKFQOVSKELIGLVLTKNKTYRVDDIDWDQNP 243
Db 270 SHKVLRSVLDPMFNFYHQTTEHKFQOVSKELIGLVLTKNKTYRVDDIDWDQNP 329

QY 244 STFKKADGSEVSLEYRYRKOYNOEITDLKQVLSQPKRRRGGTLPGPAMLIPELCYL 303
Db 330 STFKKADGSEVSLEYRYRKOYNOEITDLKQVLSQPKRRRGGTLPGPAMLIPELCYL 389

QY 304 TGLTDKMRDNFNKDLAVHTRLTPEQRQREVGRLLIDYIHKNQNVQRELWDGSLFSDNL 363
Db 390 TGLTDKMRDNFNKDLAVHTRLTPEQRQREVGRLLIDYIHKNQNVQRELWDGSLFSDNL 449

QY 364 LSFSGRILQTEKHOGGKTFDYNPQADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA 423
Db 450 LSFSGRILQTEKHOGGKTFDYNPQADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA 509

QY 424 SLIQNLFKVTPANGQMKAIMEVDRTTEAYLRVLQKQVTDATQIVVCLLSNRKDKYD 483
Db 510 SLIQNLFKVTPANGQMKAIMEVDRTTEAYLRVLQKQVTDATQIVVCLLSNRKDKYD 569

QY 484 AIKKYLCTDPTSPQCVVARTLGKQTVMAITKIALQNCXGGLWRVDIPLKLVMI 543
Db 570 AIKKYLCTDPTSPQCVVARTLGKQTVMAITKIALQNCXGGLWRVDIPLKLVMI 629

QY 544 GIDCVHDMTAGRSIAGFVASINEGTRFSCIFQDRGOELVDGLKVCLOALRAWNSC 603
Db 630 GIDCVHDMTAGRSIAGFVASINEGTRFSCIFQDRGOELVDGLKVCLOALRAWNSC 689

QY 604 NEYMPRIIVYRDGVDGGLKTLVNVVEPQFLDCLKSIGRGNPRILTVLVKKRVNTRFF 663
Db 690 NEYMPRIIVYRDGVDGGLKTLVNVVEPQFLDCLKSIGRGNPRILTVLVKKRVNTRFF 749

QY 664 AQSGRLQNLPLGTVIDVETRPEDWDFPIVQAVRSQSVSPTHYNYDNSGLKPDHIQ 723
Db 750 AQSGRLQNLPLGTVIDVETRPEDWDFPIVQAVRSQSVSPTHYNYDNSGLKPDHIQ 809

QY 724 RLTYKLCIIYNNPQVIRVPAPQVAHKLAFVVGSIHREPNIISNRLIYL 775
Db 810 RLTYKLCIIYNNPQVIRVPAPQVAHKLAFVVGSIHREPNIISNRLIYL 861

RESULT 4
ID O95404 PRELIMINARY; PRT; 861 AA.
AC O95404;

```

```

DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE HIWI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99069219; PubMed=9851978;
RA Cox D.N., Chao A., Baker J., Chang L., Qiao D., Lin H.;
RT "A novel class of evolutionarily conserved genes defined by piwi are
essential for stem cell self-renewal.";
RL Genes Dev. 12:3715-3727 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Qiao D., Zeeman A.-M., Deng W., Looijenga L.H.J., Lin H.;
RT "Molecular characterization of hiwi, a human member of the piwi stem
cell gene family whose overexpression is correlated to seminomas.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF104260; AAC97371.2; -.
DR Genew; HGNC:9007; PIWIL1.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PSS0821; PAZ; 1.
DR PROSITE; PSS0822; PIWI; 1.
SQ SEQUENCE 861 AA; 98530 MW; 43D7F60E99D997B7 CRC64;

Query Match 99.2%; Score 4071; DB 4; Length 861;
Best Local Similarity 99.7%; Pred. No. 1e-315;
Matches 770; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GVNTRQNLHVKESKTGSSGIIIVRLSTNNHRLTSRQWALYQYHIDYNPMEARRLSAL 63
Db 90 GVNTRQNLHVKESKTGSSGIIIVRLSTNNHRLTSRQWALYQYHIDYNPMEARRLSAL 149

QY 64 LFOHEDLIGKCHAFDGTILFLPKRLQKQVTEVFSKTRNGEDVRITITLTNELPPTSPCTL 123
Db 150 LFOHEDLIGKCHAFDGTILFLPKRLQKQVTEVFSKTRNGEDVRITITLTNELPPTSPCTL 209

QY 124 QFYNIIFRLLKIMNLQOIGRNYNPNPIDIPSHRLVWPFGTTSIIQYENSIMLCTDV 183
Db 210 QFYNIIFRLLKIMNLQOIGRNYNPNPIDIPSHRLVWPFGTTSIIQYENSIMLCTDV 269

QY 184 SHKVLRSVLDPMFNFYHQTTEHKFQOVSKELIGLVLTKNKTYRVDDIDWDQNP 243
Db 270 SHKVLRSVLDPMFNFYHQTTEHKFQOVSKELIGLVLTKNKTYRVDDIDWDQNP 329

QY 244 STFKKADGSEVSLEYRYRKOYNOEITDLKQVLSQPKRRRGGTLPGPAMLIPELCYL 303
Db 330 STFKKADGSEVSLEYRYRKOYNOEITDLKQVLSQPKRRRGGTLPGPAMLIPELCYL 389

QY 304 TGLTDKMRDNFNKDLAVHTRLTPEQRQREVGRLLIDYIHKNQNVQRELWDGSLFSDNL 363
Db 390 TGLTDKMRDNFNKDLAVHTRLTPEQRQREVGRLLIDYIHKNQNVQRELWDGSLFSDNL 449

QY 364 LSFSGRILQTEKHOGGKTFDYNPQADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA 423
Db 450 LSFSGRILQTEKHOGGKTFDYNPQADWSKETRGAPLISVKPLDNWLLIYTRRNYEAA 509

QY 424 SLIQNLFKVTPANGQMKAIMEVDRTTEAYLRVLQKQVTDATQIVVCLLSNRKDKYD 483
Db 510 SLIQNLFKVTPANGQMKAIMEVDRTTEAYLRVLQKQVTDATQIVVCLLSNRKDKYD 569

QY 484 AIKKYLCTDPTSPQCVVARTLGKQTVMAITKIALQNCXGGLWRVDIPLKLVMI 543
Db 570 AIKKYLCTDPTSPQCVVARTLGKQTVMAITKIALQNCXGGLWRVDIPLKLVMI 629

```

```

QY 544 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFODRGQELVDGLKVCLOALRAWNSC 603
DB 630 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFODRGQELVDGLKVCLOALRAWNSC 689
QY 604 NEYMPRIIVYRDGVGDGQLKTLVNVYEVPOFLDCLKSIGRGYNPRLTVIVVKRVNTRFF 663
DB 690 NEYMPRIIVYRDGVGDGQLKTLVNVYEVPOFLDCLKSIGRGYNPRLTVIVVKRVNTRFF 749
QY 664 AQSGGLQNLPLPGTVIDVEVTRPEWYDFIVSQVRSVSPHYNNVYDNSGLKPDHIQ 723
DB 750 AQSGGLQNLPLPGTVIDVEVTRPEWYDFIVSQVRSVSPHYNNVYDNSGLKPDHIQ 809
QY 724 RLTYKLCHYYNWPVIRVPAPCOYAHKLAFLVGSIHREPNLSNRLYYL 775
DB 810 RLTYKLCHYYNWPVIRVPAPCOYAHKLAFLVGSIHREPNLSNRLYYL 861

RESULT 5
QJMB7
ID Q9JMB7 PRELIMINARY; PRT; 862 AA.
AC Q9JMB7
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE MIWI (piwi).
GN PIWILL OR MIWI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyagawa S.K., Kimura T., Nakano T.;
RT "Molecular Cloning and characterization of Piwi family genes.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Deng W., Lin H.;
RT "Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
RT essential for spermatogenesis.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB032604; BAA93705.1; -
DR EMBL; AF438405; AAL31014.1; -
DR MGD; MGI:1928897; Piwi1.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
SQ SEQUENCE 862 AA; 98574 MW; 45588D13284CCC4C CRC64;

Query Match 97.0%; Score 3980; DB 11; Length 862;
Best Local Similarity 96.4%; Pred. No. 1.8e-308;
Matches 744; Conservative 19; Mismatches 9; Indels 0; Gaps 0;

QY 4 GYNTQNLHDVKESTGSSGIIVRLTSRPNALYQYHIDYNPLMEARLSAL 63
DB 91 GYNTQNLHDVKESTGSSGIIVRLTSRPNALYQYHIDYNPLMEARLSAL 150
QY 64 LFQHDGLKCHAFDGTILFLPKRLOQKVTYFSTRNGEDVRIITLTNELPPTSPCL 123
DB 151 LFQHDGLKCHAFDGTILFLPKRLOQKVTYFSTRNGEDVRIITLTNELPPTSPCL 210
QY 124 QFYNIIFRLLKIMNLQIGRNYNPNPDIDIPSHRLVWPQFTTSIIQYENSIMLCTDV 183
DB 211 QFYNIIFRLLKIMNLQIGRNYNPNPDIDIPSHRLVWPQFTTSIIQYENSIMLCTDV 270
QY 184 SHKVLRSVTLDFMNFYHQTEHKFQGVSKELIGLVLTLYNNKTRVDDIDWDQNPX 243
DB 271 SHKVLRSVTLDFMNFYHQTEHKFQGVSKELIGLVLTLYNNKTRVDDIDWDQNPX 330

```

```

QY 244 STFKKADSGSEVSFLBYRKQYNQEIITDLKQPVLSQPKRRRPGGTLPGFAMLIPELCYL 303
DB 331 STFKKADSGSEVSFLBYRKQYNQEIITDLKQPVLSQPKRRRPGGTLPGFAMLIPELCYL 390
QY 304 TGLTDKMRDNFNWMDLAVHTRLTPEQREVRGLLDYIHKNQDNVQRELRDGLSFDNSL 363
DB 391 TGLTDKMRDNFNWMDLAVHTRLTPEQREVRGLLDYIHKNQDNVQRELRDGLSFDNSL 450
QY 364 LSFSGRIILQTEKHOGGKTFDYNPQFADWSKETRCAPLISVAPLQNWLLIYTRRYEAAAN 423
DB 451 LSFSGRIILQSEKHOGGKTFDYNPQFADWSKETRCAPLISVAPLQNWLLIYTRRYEAAAN 510
QY 424 SLIQNLKFKVTPAMGNOMKKAIMEVDDRETEAYLRVLOQKVTDATQIVVCLLSNRRKDYD 483
DB 511 SLIQNLKFKVTPAMGIQMKKAIMEVDDRETEAYLRVLOQKVTSDTQIVVCLLSNRRKDYD 570
QY 484 AIKKYLCWDCPTSPSCVAVARTILGKQOTVMAATKIALQNCXMGGLWRVDIPLKLMIV 543
DB 571 AIKKYLCWDCPTSPSCVAVARTILGKQOTVMAATKIALQNCXMGGLWRVDIPLKLMIV 630
QY 544 GIDCYHDMTAGRRSIAGFVASINEGTRWFSRCIFODRGQELVDGLKVCLOALRAWNSC 603
DB 631 GIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFODRGQELVDGLKVCLOALRAWNSC 690
QY 604 NEYMPRIIVYRDGVGDGQLKTLVNVYEVPOFLDCLKSIGRGYNPRLTVIVVKRVNTRFF 663
DB 691 NEYMPRIIVYRDGVGDGQLKTLVNVYEVPOFLDCLKSIGRGYNPRLTVIVVKRVNTRFF 750
QY 664 AQSGGLQNLPLPGTVIDVEVTRPEWYDFIVSQVRSVSPHYNNVYDNSGLKPDHIQ 723
DB 751 AQSGGLQNLPLPGTVIDVEVTRPEWYDFIVSQVRSVSPHYNNVYDNSGLKPDHIQ 810
QY 724 RLTYKLCHYYNWPVIRVPAPCOYAHKLAFLVGSIHREPNLSNRLYYL 775
DB 811 RLTYKLCHYYNWPVIRVPAPCOYAHKLAFLVGSIHREPNLSNRLYYL 862

RESULT 6
QJNA60
ID QJNA60 PRELIMINARY; PRT; 829 AA.
AC QJNA60;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ35814 (piwi).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahori K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK093133; BAC04068.1; -
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS50821; PAZ; 1.
DR PROSITE; PS50822; PIWI; 1.
KW Hypothetical protein.
SQ SEQUENCE 829 AA; 94816 MW; 3AF9FD1D0FA9F5A7 CRC64;

Query Match 94.7%; Score 3883; DB 4; Length 829;
Best Local Similarity 99.9%; Pred. No. 9.5e-301;

```


01-MAR-2003 (TREMELrel..23, Last annotation update)
DE Similar to piwi like homolog 1 (Drosophila).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strauberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC031060.1; -
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0821; PAZ; 1.
DR PROSITE; PS0822; PIWI; 1.
SQ SEQUENCE 852 AA; 96588 MW; 37769E078B96D13 CRC64;

Query Match 51.3%; Score 2103.5; DB 4; Length 852;
Best Local Similarity 51.2%; Pred. No. 9.3e-159;
Matches 398; Conservative 146; Mismatches 225; Indels 9; Gaps 4;
QY 1 MIFGVNTRQNLHDVKESTGSSGIIVLSTNHFRLTSPQWALYQYHIDYNPLMEARLR 60
DB 81 MDLSICTREKLAHVRNCKTGSSGIPVKLVNLFNLPDQWQLYQYHVTIPDLASRLR 140
QY 61 SALLFOHEDLIGKCHAFDGTILFLPKRLOQKTEVFSKTRNGEDVRITITLNLPTSP 120
DB 141 IALLYSHSELNSKAKAFDGAIFLSQLEKTELSSETORGETIKMTILKRELPSGP 200
QY 121 TLOQYNIIFRLLKIMNLQOIGRNYNPNDDIPSHRLVWPQFTTSILQYENSIMLC 180
DB 201 VCIOQVNIIFRLLKIMNLQOIGRNYNPNDDIPSHRLVWPQFTTSILQYENSIMLC 260
QY 181 TDVSHKVLRSSETVLDPMFNFYHOTBEHKFOEVSKEKELIGLWLTKNKTYRVDIDWDQ 240
DB 261 ADVSKVLRNETVLEFMTALCORTGLSCFTQCEKQLIGLIVLTRYNNRTYSIDIDWSV 320
QY 241 NPKSTFKKADGSEVSFLEYRYKQYNOEITDKQPLVLSQPKRRRFGGTLPGPAMLIPEL 300
DB 321 KPHTFTQKRDGTETIYVYKQYDIYVSNLQPMVLSLKKRNDNSEAQ-LAHLIPEL 379
QY 301 CYLTGLTDQATSDQFQMKAVAEKTRLSQGRQRLARLVNDIQRNTNAREFELETWGLHFG 439
QY 361 SNLLSFGSRILQTEKHOGGKTFDYNPO---FADWSKETRGAPLISVKPLDNWLLIYTRR 417
DB 440 SQ-1SLTGRIVPSEKILMQ---DHICQPVSAADWSKDRTCKILNAQSLNTWLLCSDR 494
QY 418 NYEAANSILQNLFPKVPAMGMQMKAIMIEVDDETEAYLRVLOQKVTADTQIVVCLLSSN 477
DB 495 TEYVAESPLNCLRRVAGSMGFNDVPKIKIQENPAAFPVRAIQYVDVQVLCILPSN 554
QY 478 RKDYADALKYLCCTCPSPSCVARTLGKQQTWAIATKALQNCQKMGELNRVDIPL 537
DB 555 QRTYDSIKKYSDDCPVSPSCVARTLGKQQTWAIATKALQNCQKMGELNRVDIPL 614
QY 538 KLVMIYDIDYHDMTAGRRSTAGFVASINEGRTWFSRCIFQDRGOELVDGLKVCLOAAL 597
DB 615 KSLAVVGIDVCKDALSKDVVGVGVSNPRTWFSRCILQRTWIDVADCLKVFMTGAL 674
QY 598 RAMWSCNEMPSRIIVYRGVGDGQKTLVNYEVPQFLDCLKISIRGYNPLTIVVVKR 657
DB 675 NKWYKXNDLPARIIVYRAGVGDGQKTLVNYEVPQFLDCLKISIRGYNPLTIVVVKR 734
QY 658 VNTFFAQSGLRQNLPGTVIDVEVTPENYDFIVSQVRSGSVGPTHYNTYDNSGL 717
DB 735 CMPREFTEMNRVQNPPLGTIVVDSEATRNENYDFILISQVACRGTVSPTYNYVDNGL 794
QY 718 KPDHQRITLYKLCYIYNNWPGVIRPAPCOYAHKLAFLVGOSIHRBNLSNRLYL 775

DB 795 KPDHQRITLYKLCYIYNNWPGVIRPAPCOYAHKLAFLVGOSIHRBNLSNRLYL 852

RESULT 9
Q8N9V8

ID Q8N9V8 PRELIMINARY; PRT; 852 AA.
AC Q8N9V8;
DT 01-OCT-2002 (TREMELrel..22, Created)
DT 01-OCT-2002 (TREMELrel..22, Last sequence update)
DT 01-MAR-2003 (TREMELrel..23, Last annotation update)
DE Hypothetical protein FLJ36156 (piwi).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
EL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK093475; BAC04179.1; -
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0821; PAZ; 1.
DR PROSITE; PS0822; PIWI; 1.
KW Hypothetical protein.
SQ SEQUENCE 852 AA; 96560 MW; 75C6EFA4E70701B CRC64;

Query Match 51.2%; Score 2101.5; DB 4; Length 852;
Best Local Similarity 51.0%; Pred. No. 1.3e-158;
Matches 397; Conservative 148; Mismatches 224; Indels 9; Gaps 4;
QY 1 MIFGVNTRQNLHDVKESTGSSGIIVLSTNHFRLTSPQWALYQYHIDYNPLMEARLR 60
DB 81 MDLSICTREKLAHVRNCKTGSSGIPVKLVNLFNLPDQWQLYQYHVTIPDLASRLR 140
QY 61 SALLFOHEDLIGKCHAFDGTILFLPKRLOQKTEVFSKTRNGEDVRITITLNLPTSP 120
DB 141 IALLYSHSELNSKAKAFDGAIFLSQLEKTELSSETORGETIKMTILKRELPSGP 200
QY 121 TLOQYNIIFRLLKIMNLQOIGRNYNPNDDIPSHRLVWPQFTTSILQYENSIMLC 180
DB 201 VCIOQVNIIFRLLKIMNLQOIGRNYNPNDDIPSHRLVWPQFTTSILQYENSIMLC 260
QY 181 TDVSHKVLRSSETVLDPMFNFYHOTBEHKFOEVSKEKELIGLWLTKNKTYRVDIDWDQ 240
DB 261 ADVSKVLRNETVLEFMTALCORTGLSCFTQCEKQLIGLIVLTRYNNRTYSIDIDWSV 320
QY 241 NPKSTFKKADGSEVSFLEYRYKQYNOEITDKQPLVLSQPKRRRFGGTLPGPAMLIPEL 300
DB 321 KPHTFTQKRDGTETIYVYKQYDIYVSNLQPMVLSLKKRNDNSEAQ-LAHLIPEL 379
QY 301 CYLTGLTDQATSDQFQMKAVAEKTRLSQGRQRLARLVNDIQRNTNAREFELETWGLHFG 439
QY 361 SNLLSFGSRILQTEKHOGGKTFDYNPO---FADWSKETRGAPLISVKPLDNWLLIYTRR 417
DB 440 SQ-1SLTGRIVPSEKILMQ---DHICQPVSAADWSKDRTCKILNAQSLNTWLLCSDR 494
QY 418 NYEAANSILQNLFPKVPAMGMQMKAIMIEVDDETEAYLRVLOQKVTADTQIVVCLLSSN 477

Db 495 TEYVAESFLNCLRRVAGSMGFNVDPYKIKVOENPAAFVRAIQYVDPDQVLMCILPSN 554
 Qy 478 RDKYDAIKKYLCTDPCPSQCVARTLGGKQVMAIAIKALQNCCKGGLWRVDPIL 537
 Db 555 QTYYSIIKIISSDCPVSCVCLARTLNKQNMMSIAIKAMQTKGGLWAVEIPL 614
 Qy 538 KLVMIVGIDCYHDMTAGRSIAGFVASINEGTRWFSRCIFQDRGQELVDGLKVCLOAAL 597
 Db 615 KSLWVGIDVCKDALSXDVMVGVCAVNPRIITRWFSCILORTMTDADCLKVFMTGAL 674
 Qy 598 RAWNSCNEWPSRIIVYRDGVDGOLKTLVNYEVPQFLDCLKSIIGRGYNPRLTVIVVKR 657
 Db 675 NKWYKINHDLPARIIYIRAGVGGQLETLIEVEVQLLSVAESGTSLSLSVIVVRK 734
 Qy 658 VNTREFAGSGRLQPLPGTVIDVEVTRPEWYDFIVSQVRSVSPHYNYIYDNSGL 717
 Db 735 CNPRFFTEMNRTVQNPPLGTVDSEATRNWDYFLISQVACRGTVSPYYNVIYDNGL 794
 Qy 718 KPDHQRILTYKLCHYYNWPVIRVPAPCOYAHKLAFLVQGSIHREPILSNRLYYL 775
 Db 795 KPDHQRILTYKLCHYYNWPVIRVPAPCOYAHKLAFLVQGSIHREPILSNRLYYL 852

RESULT 10

Q9GPA8 ID Q9GPA8 PRELIMINARY; PRT; 854 AA.
 AC Q9GPA8;
 DT 01-MAR-2001 (TREMELrel. 16, Created)
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
 DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
 DE Seawi.
 OS Strongylocentrotus purpuratus (Purple sea urchin).
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
 OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
 OC Strongylocentrotus.
 NCBI_TaxID=7668;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rodriguez A.J., Sonder E.M.;
 RT "Seawi - Cloning and Molecular Characterization of a Sea Urchin
 RT Homolog of Pw1."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY014899; AAG42533.1; --
 DR InterPro; IPR003100; PAZ.
 DR InterPro; IPR003165; Piwi.
 DR Pfam; PF02170; PAZ; 1.
 DR Pfam; PF02171; Piwi; 1.
 DR PROSITE; P50821; PAZ; 1.
 DR PROSITE; P50822; PIWI; 1.
 SQ SEQUENCE 854 AA; 96722 MW; 509A1D39C0D1922C CRC64;

Query Match 49.0%; Score 2009.5; DB 5; Length 854;
 Best Local Similarity 50.5%; Pred. No. 2.9e-151;
 Matches 386; Conservative 133; Mismatches 236; Indels 9; Gaps 6;
 Qy 14 VKESKTGSGGIIIVRLSTNHFRLTSRQWALYQVHDYNDYFPLMEARRLSALLFQHEDLICK 73
 Db 98 VKQAIAGDK---IALIANGFKLTKTDQWLYQVYRVDPEPEILNPRAPFALLGHSALLGK 154
 Qy 74 CHAFDGTILFLPKRLOQKTEVSKTRNGEDVRITITLNEIPPTSPCTCLQFYNIIFRL 133
 Db 155 GLTLDMDTVSLFKLAEKVTNLGAERKGSNNVNCVTHVATLNPAPNTLHLYNLFRR 214
 Qy 134 LKIMLQOIGRNYPNDPDIPIHRLVTPGFTTSILOVENSIMLCTDVSHKVLRSVTV 193
 Db 215 LKMINNEQVGRNYDPTAALDIKQGLQWPGFVTSILOVEDVWLLSDISHKVLRTQV 274
 Qy 194 LDFNFNYHQTEHFKQEQVSKELIGLVLTLYNNKNTYRVDIDMDQNPKSTFKKADGSE 253
 Db 275 WEVANDLFNKA-RGRPKKEITKLWIGQVLTLYNNKNTYRVDIDMDQNPKSTFKKADGSE-P 332
 Qy 254 VSLEYRYKQYNGEITDLKQPVLYSQKRRGPGGTLPGAMLIPLCYLTGLTGMND 313

Db 333 VSTVDYFKSYERYIHDVNPMLVSRPKREKKGV--GPAYLPPELCELTLGLSDMDRAD 390
 Qy 314 FNVKDLAVHTLTPBOROREVRLIDYTHKNDVQRELDRDGLSFDLSNLSFSGRILQT 373
 Db 391 FNVKDMGQHTRVGPDRCTLSGFIKKLSNEEVKTYLDSGMEFDEKQVLTGEVLEPP 450
 Qy 374 EKHQGGKTFDYNPOPADWSKETRGAPLISVLPKLDNWLIIYTRRYEAAANSIQLFKVT 433
 Db 451 EKLFRSGKQFSYNPSNADWSRDRTRGNALTDKILANNKIFYTRRDANRGQDFIKSLRVA 510
 Qy 434 PANGMOMKAIWLV--DDRTEAVRLVLOQVADTQIVVCLLSSNRKDYDAIKYLCTD 492
 Db 511 NPMGMNVGRPEIVELPDDRTIETYSLOAQIAQDTQIVVILPTNRKDYDAIKKTCVVT 570
 Qy 493 CPTPSQCVVARTLGGKQVMAIAIKALQNCCKGGLWRVDPILKLMVIGIDCYHDMT 552
 Db 571 HPCPSQVIVSRTLKQOMLMSVATKIAMQNCCKGGLWRVDPILKLMVIGIDCYHDMT 630
 Qy 553 AGERSIAGFVASINEGTRWFSRCIFQDRGQELVDGLKVCLOAALRAWNSCNEWPSRII 612
 Db 631 TKGRSVLGFVSNMKNKSTQTSFSSCAPQHAQGFGANLSTLMNNALKRYTQINEKPFERI 690
 Qy 613 VYRDGVGDGOLKTLVNYEVPQFLDCL-KSIGRGYNPRLTVIVVKKRVNTRFFAQSGRLQ 671
 Db 691 IFEDGVGDSQVNLVDYELKQIKDLYVPGTVHKLAVVVVKKKINNRFANLGGLS 750
 Qy 672 NPLPGTVIDVEVTRPEWYDFIVSQVRSVSPHYNYIYDNSGLKPDHIOBLTYKLCH 731
 Db 751 NPPEGTVIDVWTKPHLYDYFIISQVRSVSPSYNVVYDTTGLKPDHIOBLTYKLCH 810
 Qy 732 IYNNWPGVIRVPAPCOYAHKLAFLVQGSIHREPILSNRLYYL 775
 Db 811 LYFNWPGTVIRVPAPCOYAHKLAFLVQGSIHREPILSNRLYYL 854

RESULT 11
 Q8CGT6 ID Q8CGT6 PRELIMINARY; PRT; 878 AA.
 AC Q8CGT6;
 DT 01-MAR-2003 (TREMELrel. 23, Created)
 DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE Miwi 2 protein.
 GN MIW12.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Carmell M.A., Xuan Z., Zhang M.Q., Hannon G.J.;
 RT "The Argonaute family: tentacles that reach into RNAi, developmental
 RT control, stem cell maintenance, and tumorigenesis."
 RL Genes Dev. 16:2733-2742(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Adams M., Mural R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY135692; AAN75583.1; -- 3C6265E0E4ED7C3F CRC64;
 SQ SEQUENCE 878 AA; 98617 MW; 3C6265E0E4ED7C3F CRC64;

Query Match 45.3%; Score 1859.5; DB 11; Length 878;
 Best Local Similarity 46.5%; Pred. No. 2.8e-139;
 Matches 370; Conservative 134; Mismatches 209; Indels 83; Gaps 10;
 Qy 4 GVNTRQMLDHWKSKTGSSGIIIVRLSTNHFRLTSRQWALYQVHDYNDYFPLMEARRLSAL 63
 Db 142 GVCRTREKLTWKDCKTGSSGIPVRLVTLNLFNLDLPQDWLYQVHYVTSPLASRRRLAL 201
 Qy 64 LFOHEDLIGHCHAFDGTILFLPKRLOQKTEVSKTRNGEDVRITITLNEIPPTSPCTCL 123
 Db 202 LYNHSLSDKAKAPDAGSLFLSEKLDQKVTLTSTQRTGRTIKITLITLTKLFPNFPVCI 261

```

QY 124 QFVNIIFRELLKIMNLQOIGRNYNPNPDIPIPSHRLVWPGFTTSLIQYENSIMLCITDV 183
DB 262 QFNVIFRILKMLSNYQIGRNYFKESEVEIP-----QY-NKLFNADV 305
QY 184 SHKVLASETVLDFPMFNYHQTEHKEFQEQVSKELIGLVLTLYNNKTYRVDIDWDQNP 243
DB 306 NYKVLRSNETVDFMTDLCLRTGNSCFTECHKQGLVGLVLTLYNNKTYRIDDDWSVKPT 365
QY 244 STFKKADGSEVSELYRKYQYQVBEITDLKQPVLSQPKRRRPGGTLPGPAMLIPELCYL 303
DB 366 QAFQKDGSEVTVYDYKQYDITLSDNQPVLSLKKRNDSE-PQWHLMPLELCFL 424
QY 304 TGLTDKMRDNFMKDLAVHRLTPQROREYGRLDYIHKNDNVQRELRD-----W-- 355
DB 425 TGLSSQATSDFLMKAVAEETRLSPVGRQQQLARLVDDIORTLPSSQEVLSHTSLPLAP 484
QY 356 ---GLSFD---SNLSFSGRILQTEKHOG-----GKTFDYNPQF-ADWSKETRGA 399
DB 485 EPGGLSAPLSTVLPFAOQLLTALSFGIPLPHLKPSFPLFCQAPAFADWSKDMRSC 544
QY 400 ELISVPLDNWLLIYTRNRYEAANSILQNLKVTYPAMGMQKAMIEVDRTTEAYLRVL 459
DB 545 KVLSSQFLRWLVCCNRAHLEIAEPLSCLRRVGGSMGNVGPYKLIIVDETPAAPFLRAI 604
QY 460 QOKVTADTQIVVCLLSSNRDKYDAIKYLCTDCTPTSCQVAVRTLGKQOQVMAIATKIA 519
DB 605 QVHGDEPDLVLCILPSNQYVDSIKYLSLSDCPVPSQCVLTRLNKQGTMLSVATKIA 664
QY 520 LQWNCRMGELMRYDTPKLVMIVGIDCHYDMTAGRRSAGFVASINEGMRWFSRCIFQ 579
DB 665 MQNTCKLGGELMSVEIPLSLMVGDI CRDALNKNVWVGVFASINSRITRWFSEKVLQ 724
QY 580 DRQOELVDGLKVCLOAALRAWNSCNEYPMSRIIVYRDGVDGOLKTLVNYEVPQSLDCLK 639
DB 725 RTAADIADCLKCMTCALNRWYHNDLPARIVYRDGVDGOLKAVLEYEVPQLKSVT 784
QY 640 SIGRGNPLRTIVVKKRVNTRFPAQSGRLQNLPGTVIDVEVTRPEWYDFIVSQVR 699
DB 785 ECG-----SDAR-----YDFYLSOTAN 802
QY 700 SGVSPTHYVNIYDNLGKPDHQRITLYKLYNNWPGVIRVPAPCOYAHKLAFLVNGS 759
DB 803 RGTVSTHYNIVYDNLALPDMQRITFLKLYNNWQGLISVPAPCOYAHKLTFLVAQS 862
QY 760 IHREPNSLSNRLYYL 775
DB 863 VHKEPSLELANNLFYL 878

```

RESULT 12

```

QSN8G9 PRELIMINARY; PRT; 666 AA.
AC QSN8G9;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ39518 (piwi).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Eutheria; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,
RA Wakebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Magatsuma M., Murakawa K., Kanekori Y., Takahashi-Fujii A., Oshima A.,
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K.,
RA Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project."

```

Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

EMBL: AK096377; BAC04873.1; --
 InterPro: IPR003100; PAF;
 InterPro: IPR003165; PAF;
 Pfam: PF02170; PAF; 1.
 Pfam: PF02171; PAF; 1.
 PROSITE: PS0821; PAF; 1.
 PROSITE: PS0822; PAF; 1.
 Hypothetical protein.
 QY SEQUENCE 666 AA; 76099 MW; 1082D86916390FA8 CRC64;

Query Match 44.8%; Score 1838.5; DB 4; Length 666;
 Best Local Similarity 51.5%; Pred. No. 8.8e-138;
 Matches 346; Conservative 127; Mismatches 190; Indels 9; Gaps 4;

```

QY 107 ITITLNEPPTSPCTCLOFYNIIFRLKIMLQOIGRNYNPNPDIPIPSHRLVWPGF 166
DB 1 MTITLRELPSPPVCIOVFNIFRKLKLSYQIGRNYNPNPDIPIPSHRLVWPGF 60
QY 167 TTSILOVENSIMLCTDVSHKVLRSSETVDFMNFVHOTEHKEFQEQVSKELIGLVLTXY 226
DB 61 AISVSIFERKLIFSADVSXKVLNRTVLEFMTALCOETGLSCFTQTCERKQLGLIVLTY 120
QY 227 NNTYVDDIDWDQNPSTFKKADGSEVSELYRKYQYQVBEITDLKQPVLSQPKRRRPG 286
DB 121 NNTYVDDIDWDQNPSTFKKADGSEVSELYRKYQYQVBEITDLKQPVLSQPKRRRPG 180
QY 287 GGTLPAPMLIPELCYLCTGLTDKMRDNFMKDLAVHRLTPQROREYGRLDYIHKND 346
DB 181 NSB-PQALHILPELCFLTGITDQATSDFLMKAVAEETRLSPVGRQQQLARLVDDI 239
QY 347 NVQRELDRGLSDPSNLLSFSGRILQTEKHOGKTFDYNPQ---FADWSKETRGAFLIS 403
DB 240 NARFELETWGLHFGSQ-ISLTGRIVSEKILMQ---DHICQPVSAADWSKDRTCKILN 294
QY 404 VKPDLNWLIIYTRNRYEAANSILQNLKVTYPAMGMQKAMIEVDRTTEAYLRVLOKV 463
DB 295 AQSINTMLICSDRTEYVAESFLNCLRRVTGSGFNVDYPKIKVQENPAARFVRAIQTV 354
QY 464 TADTQIVVCLLSSNRDKYDAIKYLCTDCTPTSCQVAVRTLGKQOQVMAIATKIALQMN 523
DB 355 DPVQVLMCLPSNQYVDSIKYLSLSDCPVPSQCVLTRLNKQGTMLSVATKIAMQMT 414
QY 524 CRGGELMRYDTPKLVMIVGIDCHYDMTAGRRSAGFVASINEGMRWFSRCIFQDRQ 583
DB 415 CKLGELMRYDTPKLVMIVGIDCHYDMTAGRRSAGFVASINEGMRWFSRCIFQDRQ 474
QY 584 ELVDGLKVCLOAALRAWNSCNEYPMSRIIVYRDGVDGOLKTLVNYEVPQSLDCLKS 643
DB 475 DVADCLXVFTGALNKYKYNHDLPARIVYRDGVDGOLKTLVNYEVPQSLDCLKS 534
QY 644 GYNPRTIVVKKRVNTRFPAQSGRLQNLPGTVIDVEVTRPEWYDFIVSQVRSGSV 703
DB 535 NTSRLSIVVKKRVNTRFPAQSGRLQNLPGTVIDVEVTRPEWYDFIVSQVRSGSV 594
QY 704 SPHYNNVYDNLGKPDHQRITLYKLYNNWPGVIRVPAPCOYAHKLAFLVNGSIRHE 763
DB 595 SPHYNNVYDNLGKPDHQRITLYKLYNNWPGVIRVPAPCOYAHKLAFLVNGSIRHE 654
QY 764 PNLGSLNRLYYL 775
DB 655 PSLANLHLYL 666

```

RESULT 13

```

Q9GPA7 PRELIMINARY; PRT; 808 AA.
ID Q9GPA7
AC Q9GPA7;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE Seawi (Fragment).
OS Strongylocentrotus purpuratus (Purple sea urchin).

```

OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RA Rodriguez A.J., Bonder E.M.;
RT "Seawi - Cloning and Molecular Characterization of a Sea Urchin
Homolog of Pw1";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RM EMBL: AV014500; AGA2534.1; -
DR InterPro: IPR003100; PAZ;
DR InterPro: IPR003165; Pw1.
DR Pfam: PF02170; PAZ; 1.
DR Pfam: PF02171; Pw1; 1.
DR PROSITE: PSS0821; PAZ; 1.
DR PROSITE: PSS0822; Pw1; 1.
FT NON TER 808
SQ SEQUENCE 808 AA; 91366 MW; 19528F4B9D10474C CRC64;

Query Match 44.2%; Score 1811.5; DB 5; Length 808;
Best Local Similarity 49.4%; Pred. No. 1.7e-135;
Matches 354; Conservative 125; Mismatches 228; Indels 9; Gaps 6;

QY 14 VKESKSGSIIIVLSTNHFLTSRPOWALYOHIDYNPLMEARLRSLALFQHEDLIGK 73
DB 98 VKQAIAGDK---TALTANGFKLTKEDWQYQYVDFEPEILNPRARFALLXGHSALLGK 154

QY 74 CHAPDTEILPKRLOOKVTEVFSKTRNGEDVRITITLNLPTPTSCLOFYNIIFRL 133
DB 155 GLTLDMDTMYSLKFAEKVTLNLSAERKDGSNVICVTHVATLNPAPNTLHLVNLPRRC 214

QY 134 LKIMNLQOIGNYPNDPIDIPSHRLVTPGFTTSILOVENSIMLCTDVSHKVLSETV 193
DB 215 LKIMNEQVGRNYDDTAIDIKQHLQWPGFVTSILOVENDVLLSDISHKVLRTV 274

QY 194 LDFMFPYHQTEBHKFOEQVSKELIGLVLTLYNNKTYRVDIDMDQPKSTFKKADGSE 253
DB 275 WEVNDLFENKA-RGRFEKITLMIQIIVLTLYNNKTYRVDIDFDTPADTFERSG-P 332

QY 254 VSLEYRKOYNGEITDLKQPVLSOP-KRRPGGTLPGPAMLIPELCTGLTDMKEND 313
DB 333 VSVDVFKSYERVIHVNQPMVLSRPPKREKGV--GPAFLPELCTGLTGLSDMEAD 390

QY 314 FNVKDLAVHTLTPQORQEVGRGLIDYIHKNDNVQRELVDWGLDGLSFLSFSRIIQT 373
DB 391 FNVKDMGQHTRVGPQDRCKTTLGSGFIKLSSNEEVKTYLDSNGMEFDKEQVKLTGRVLP 450

QY 374 EKHQGGKTFDYNPOPADRSKTRGAPLISVKPLDNWLLIYTRNVEANSLIQNLFKVT 433
DB 451 EKLFORQKQFSPNSNADWSRDTGNGALTDATLNNWKIFYTRRANRGQDFIKSLRVA 510

QY 434 PAMGMQMKAIMIEV--DDRTEAYLRVLOQVTDQIVVCLSSNRKDKYDAIKKYLCTD 492
DB 511 NPMGMVVRGPEIIVELPDDRTETTSLSQAQADTQIVVVILPTNEKORYDAIKTCVVT 570

QY 493 CPTPSQCVVARTLGKQOVTMAITKALQNCMGGLMRVDIPLKLVNIGIDCYHMT 552
DB 571 HPCPSQVIVSRTLSKQOMLSVATKIAMQNCMGGLMRVDIPLKLVNIGIDCYHMT 630

QY 553 AGRSIAGFVASINEGTRWFSCIPQDRGOELVDGLKVCLOAALRAWNSCNEYWPSRII 612
DB 631 TGRSVLGFVAVNWKSTSFSSCAFOHAGGEGANLSTLMNVALKRYQVINEKEPERII 690

QY 613 VYRDGVDGQGLKTLVNYEVPQFDLCL-KSIGRGNPRLTVIVVKKRVNTRFFAQSGBLQ 671
DB 691 IFRDGVGDSQNLWDVYELKQIKTDLVKVPQGTVHKLAVVVKKVNNRFFANLRGGLS 750

QY 672 NPLPCTVIDVETRPBWYDFEIVSQAVRSQSVSTPHYNNVYDNSGLKPHRIQLTY 727
DB 751 NPPPGTVIDDVTKPHLYDYFIISQVRSQSVSTPHYNNVYDNTGLKPDHNRQLTY 806

RESULT 14
Q9JMB6 PRELIMINARY; PRT; 971 AA.
ID Q9JMB6
AC Q9JMB6;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE MILI (Miwi like).
GN PIWIL2 OR PIWIL1L OR MILI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Miyagawa S.K., Kimura T., Nakano T.;
RT "Molecular Cloning and Characterization of Piwi family Genes";
RM Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB032605; BAA93706.1; -
DR MGD; MGI:1930036; PiwiL2.
DR InterPro: IPR003100; PAZ;
DR InterPro: IPR003165; Pw1.
DR Pfam: PF02170; PAZ; 1.
DR Pfam: PF02171; Pw1; 1.
DR PROSITE: PSS0821; PAZ; 1.
DR PROSITE: PSS0822; Pw1; 1.
SQ SEQUENCE 971 AA; 109488 MW; 01E143C6513310FB CRC64;

Query Match 43.1%; Score 1766; DB 11; Length 971;
Best Local Similarity 42.9%; Pred. No. 9.5e-132;
Matches 327; Conservative 162; Mismatches 264; Indels 10; Gaps 9;

QY 18 KYGSSGIIIVLSTNHFLTSRPOWALYOHIDYNPLMEARLRSLALFQHEDLIGKCHAF 77
DB 214 KQSKGTPOSGLNLKIQCHNE-AVYQHVTFSPVCKSMRFGKMDKHOSHVGNTAF 272

QY 78 DGTILFPLRLOOKVTEVFSKTRNGEDVRITITLNLPTPTSCLOFYNIIFRLKIM 137
DB 273 DGSILYLPVKLQQVVELKSKRTDDBAISIQLTKILEPCSDLCIPFYVNVFRRVMKLL 332

QY 138 NLQOIGNYPNDPIDIPSHRLVTPGFTTSILOVENSIMLCTDVSHKVLSETVLDEM 197
DB 333 DMKLVGRNYDFPTSAVLOQHLQWPGYAAISIRTDGGLFLADVSHKVTINDSVLDM 392

QY 198 FNFYHOTEHKKFOEQVSKELIGLVLTLYNNKTYRVDIDMDQPKSTFKKADGSEVSL 257
DB 393 HAIYQNKSH-FQDECSKLLVGSIVITRYNNITYRDDVDWNTPKDSFVMSDGEITPL 451

QY 258 EYTRKOYNGEITDLKQPVLSOP-KRRPGGTLPGPAMLIPELCTGLTDMKENDNV 316
DB 452 EYYSKNYGTIVKEDDQPLIHRPSPQRNNHGMLLKGEIILLPELSFMTGIPKMKKDFRA 511

QY 317 MKDLAVHTLTPQORQEVGRGLIDYIHKNDNVQRELVDWGLSFLSFSRIIQTETKI 376
DB 512 MKDLTQOINLSFKQHHGALECLLQISQNETASNELTRWGLSHKVDHVKIEGLLPMERI 571

QY 377 HQGGKTFDYNPOPADRSKE-TRGAPLISVKPLDNWLLIYTRNVEANSLIQNLFKVTPA 435
DB 572 NLNRTSF-VTSEDLNVKVEVTRDASILTI-PMHFALFYPKRAMDQARELVNMLEKIAGP 629

QY 436 MGMQMKAIMIEV--DDRTEAYLRVLOQ--KVTDQIVVCLSSNRKDKYDAIKKYLCTD 492
DB 630 IGMRIISPPAWVELKDDRIETIYRTQSLGVGEGKIQMVVCIITMGTRDDLYGAIKKLCVQ 689

QY 493 CPTPSQCVVARTLGKQOVTMAITKALQNCMGGLMRVDIPLKLVNIGIDCYHMT 552
DB 690 SPVPSQVINVRTIGQPTLRSVAQKILLQNCMLGGLGWDVDPILKQLMVIGMDVYHDS 749

QY 553 AGRSIAGFVASINEGTRWFSCIPQDRGOELVDGLKVCLOAALRAWNSCNEYWPSRII 612
DB 750 RGRSVVGFVASINLTLTKWRSVVFQMPHOBIVDSLCLVGLSKLYKYEVNHLPEKIV 809

QY 613 VYRDGVDGQGLKTLVNYEVPQFDLCLCSIGRGNPRLTVIVVKKRVNTRFFAQSGBRLQ 672

```

Db 810 VYRDGVSQGLKTVANEYIPQKCFEAD-NYHPKVVVQKISTNLVLAAPDHPVT 868
Qy 673 PLPGTVIDVEVTRPEWYDFIVSQAVRSVSPHYNNVYDNGSLKPDHIOQLTYKLCHI 732
Db 869 PSFGTVVDHTITSCWVDFYLLAHVROCGGTPHYICVLNTPANLSPDHMQRLTFKLCHM 928
Qy 733 YNNWPGVIRVPAPCOYAHKLAFLVGSIHREPNLSISNRLYYL 775
Db 929 YNNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 971

RESULT 15
Q8CDG1 PRELIMINARY; PRT; 971 AA.
AC Q8CDG1,
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Piwi like homolog 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR ENBL; AK030116; BAC26791.1; -.
SQ SEQUENCE 971 AA; 109458 MW; B63FE11A2E9DA0F9 CRC64;

```

```

Query Match 43.0%; Score 1764; DB 11; Length 971;
Best Local Similarity 42.9%; Pred. No. 1.4e-131;
Matches 327; Conservative 162; Mismatches 264; Indels 10; Gaps 9;

Qy 18 KTSSGIIVRLSTNHPRLTSRPOWALYQYHIDYNPLMEARRLRSLALFOHEDLIGKCHAF 77
Db 214 KQSGKTPQSLGLNLKIQCHNE-AYQYHVFPSPSVCKSMRFGMLKDHQSVGTGNVAF 272
Qy 78 DGTILFLPKLQKQVTEVSKTANGEDVTRITITLWELPPTSCLOFNIIFRLLKIM 137
Db 273 DGSILYLPVKLQVWELKQSKRTDDAEISIKIQLTKILEPCSDLCIPFYNNVFRVMKIL 332
Qy 138 NLQOIGRNYNPNDDPIDPSHRLVWPFGTTSILOYENSIMLCTDVSHKVLRSSETVLDPM 197
Db 333 DMKLVGNFVDPTSAVVLQOHLQIWEFYAASIRRTDGGFLFLLADVSHKVINRDSVLDVM 392
Qy 198 FNYHOTEHKKFQEQVSKELIGLVLTKNNTYRVDDIDMQNPKSTPKKADGSEVSL 257
Db 393 HAIYQQNKER-FQDECSKLLVGSIVITRYNNRTYRDDVDNKNTPKXDSFVMSDGKEITFL 451
Qy 258 EYRKYQVNOBITLKPVLVSP-QRRRGPGTLPGLPAMLIPELCVLTGLTDMKRDENV 316
Db 452 EYISKYGITVKEDDQPLLIHRPSEQNNGMLKGEILLPELSPMGIPEAKKDFRA 511
Qy 317 MKDLAVHTRITPEQREVGRLIDYIHKNDNVORELDMGLSFDNSLLSFSGRILQTEKI 376
Db 512 MKDLTQGINLSPKQHGALCCLQRIQSONEASNELTRWGLSLHKDVHKIEGRLPMERI 571
Qy 377 HOGGKTFDYNPQADWSKE-TRGAPLISVKPLDNWLLIYTRRYEANSLLIONLFKVTFA 435
Db 572 NLRNTSF-VTSEDLNVMVKEVTRDASILTI-PMHFNLFYFKRAMDQARELVNMLEKIAGP 629
Qy 436 MGHQMKRAIMIEV-DDRTEAYLRVLOQ--KVTADTQIVVCLLSSNKDKYDAIKKYLCTD 492
Db 630 IGMRISSPAWVELKDDRIETIYTIQSLIGVEGKIQWVUCIIMGTRDDLYGAIKKLCVQ 689
Qy 493 CPTFSQCQVARTLKGQQTVMATKIALQALQNCXVGGELMRVDIPLKLVNIVGIDCYHDNT 552

```

```

Db 690 SPVPSQVINVRITIGQPTRLRSVAQKILLQMNCKLGGELAGVDIPLKQLMVGMDVHDFS 749
Qy 553 AGRSTAGFVASINEGWTFWFSRCIPQDRGQBLVDGLKVCLOALRAMNSCNEYMPSRIL 612
Db 750 RGMRSVVGFGVASTNLTLTKWYSRVVPQMPHQIIVDSLKLCLVGSLSKKYEVANHCLPEKIV 809
Qy 613 VYRDGVGDGQLKTLNNEYVPOFLDCLKSGISGRGYNPRLTIVVVKRVTNTRFFAQSGRLQN 672
Db 810 VYRDGVSDQGLKTVANEYIPQKCFEAD-NYHPKVVVQKISTNLVLAAPDHPVT 868
Qy 673 PLPGTVIDVEVTRPEWYDFIVSQAVRSVSPHYNNVYDNGSLKPDHIOQLTYKLCHI 732
Db 869 PSFGTVVDHTITSCWVDFYLLAHVROCGGTPHYICVLNTPANLSPDHMQRLTFKLCHM 928
Qy 733 YNNWPGVIRVPAPCOYAHKLAFLVGSIHREPNLSISNRLYYL 775
Db 929 YNNWPGTIRVPAPCKYAHKLAFLSGQILHHEPAIQLCGNLFFL 971

```

Search completed: December 4, 2003, 19:15:34
Job time : 45 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	1295.5	31.6	824	2	T20351	hypothetical prote	
2	1162	28.3	722	2	T30995	hypothetical prote	
3	557.5	13.6	834	2	T41568	argonaute-like pro	
4	508.5	12.4	1014	2	H86438	protein t19E23.7 [
5	505.5	12.3	891	2	T32079	hypothetical prote	
6	498.5	12.2	1000	2	T22391	hypothetical prote	
7	495.5	12.1	997	2	A84678	Argonaute (AGO)1	
8	492	12.0	988	2	T52134	Zwille protein (im	
9	488.5	11.9	930	2	A84668	Argonaute (AGO)1	
10	482	11.8	813	2	JC6569	translation initia	
11	434	10.6	887	2	T01113	translation initia	
12	429.5	10.5	1032	2	T23164	hypothetical prote	
13	429.5	10.5	1035	2	T23165	hypothetical prote	
14	421	10.3	958	2	S41013	hypothetical prote	
15	421	10.3	1040	2	D88568	protein ZK757.3 [i	
16	384	9.4	898	2	T23164	hypothetical prote	
17	343.5	8.4	1339	2	T32375	hypothetical prote	
18	316.5	7.7	905	2	T23510	hypothetical prote	
19	283	6.9	892	2	T15195	hypothetical prote	
20	280.5	6.8	945	2	T23965	hypothetical prote	
21	253	6.2	965	2	T22933	hypothetical prote	
22	252	6.1	1057	2	T16676	hypothetical prote	
23	248.5	6.1	1032	2	T19324	hypothetical prote	
24	240	5.9	1030	2	T18114	hypothetical prote	
25	238	5.8	990	2	F89325	protein T22H9.3 [i	
26	230.5	5.6	697	2	T15179	hypothetical prote	
27	225	5.5	939	2	T18974	hypothetical prote	
28	223	5.4	324	2	D87774	protein C24A11.3 [
29	220	5.4	881	2	T31818	hypothetical prote	

203 QAPIT VSEOT - OCVKT - YECVQAEWAPCVWECCIVPCSNMTNTWIVTCPCGCSCLISQV 448


```

Qy 425 LIQNLKFKVTPAMGMQMRKAIMIEV-----DDRTYLRVLQKVTADTQIVVCLLSSNRK 479
Db 450 FIEARRLGLKIQVLQGLGPMCPVIRGIPNDYLEGVKGAIKQVGDGDIHMLVWMLADNKK 509
Qy 480 DKYDAIKKYLCTDCPTPOCVVARTL-----GKQOTVMAATKIALQMNCKMGGLWR 532
Db 510 TRYDSLKFLCVCCEPIPHQCNVLRTPLAGSKDGGENKLSVLKVLQMICTKGALWKK 569
Qy 533 VDIPLKVMIVGIDCYHDMTAGRRSIAGFVASINEGMTWFSRCIFQDRGOELVDGLKVC 592
Db 570 VNIPLKNTMIYGYDLYHDSLTKGTGACVSTTSNDFQFYSQTRPHENPTQLGNNLTHF 629
Qy 593 LQAALRAWNSCNEY-WPSRIIVYRGVGDGQKTLVYVYVPOFLDCLKSIGR-----643
Db 630 VRKALKQYDSNDQFLPSRLILYRDGAGDGOIPYIKNTEVLKVRDADCAVTDKAAELSNK 689
Qy 644 -GYNPLTVIVVKKVNRTRFFAQSGRLQ---NPLPGTVIVDEVTREPWFVFFIVSQAVR 699
Db 690 VOEKIKLAFIIVTKVNMRIKQ-GSSLDNAINPQPGVVDVTVTRPERMDFILVFPQVFN 748
Qy 700 SGSVSPHYNYVNSGLKPDHIOBLTYKLCHYIYNWPGVIRVPAPCOYAHKLAFLVQGS 759
Db 749 QGTVPVSYNIHDDTDLGPKHQQLAFKLCYLYYNWQTVRVPAPCOYAHKLAFLTAQS 808
Qy 760 IHRSPNLSNRLYYL 775
Db 809 LHDDANGCLRDLPFL 824

RESULT 2
T30995
hypothetical protein C01G5.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 29-Oct-1999
C:Accession: T30995
R:Bradshaw, H.; Stellyes, L.
submitted to the EMBL Data Library, August 1999
A:Description: The sequence of C. elegans cosmid C01G5.
A:Reference number: Z20956
A:Accession: T30995
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-722 <BRA>
A:Cross-references: EMBL:U50068; PIDN:AAB37734.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Map position: IV
A:Introns: 58/3; 121/3; 202/3; 350/1; 471/3; 530/1; 609/2; 671/3
A:Note: C01G5.2

Query Match 28.3%; Score 1162; DB 2; Length 722;
Best Local Similarity 33.6%; Pred. No. 2.7e-76;
Matches 256; Conservative 154; Mismatches 279; Indels 74; Gaps 14;

Qy 41 WALYQVHYDYNPLMEARLRALLFOHEDL--IGKCHAFDGTILFLPKRLOOKTEVFSK 98
Db 6 YSIYQYHVEFPTVDEKATRENLRQPSVTVEIGKHFVFDGMILYKKEEWDQNMIEVQH 65
Qy 99 TRNGEDVRITITLNLPTSPCTCLQFVNIIFRLLKIMNQQIGRNYNENDPIDPSH 158
Db 66 PNDNSLICIRFKINRFLVDDPQINFINFIRSFDMKLTQLGRNFD-----115
Qy 159 RLVTWPQFTTSILOYSIMLCTDVSHKVRSETVDFMNFYHQTBKHFQEQVSKELI 218
Db 116 -----WDNSRAL-----RKELRS-----CONNRQRVQEKQNEVYG 145
Qy 219 GLVLTLYKANKTYVDDIDWDQNPFSFKKADGSEVSEFLYRQYQNOEITDLKQPLVS 278
Db 146 GSTITLNNKLRFTLDNEITLSFKQK-DGQIILKEYFKNQYDIDITDDQFIIS 204
Qy 279 QPKRRRGFGGTLPGFAMLIPELCVLTGLTQWRNDENVMDLAVHTRLTPEQOREVGR 338
Db 205 ESKPKQ--PGEPPQVNIYVPELCPFTGLTDMRKDFKQMKELAKHTRMSPOQRLDTRKL 262

```

```

Qy 339 IDYIHKNDVQRELRDWGLSFDLSNLSFSGRILQTEKIHOQKTFDYNPOFADSKETR 398
Db 263 ITKLSQNOTMECFQYWGISIGQDLANQARVLKSEPL-QGKR--QYEGKQAEWARGVKE 319
Qy 399 APLISVPLONWLLIYTRRYEA--ANSLIQNLFPKVTTPAMGMQMRKAIMIEV-----DDR 451
Db 320 CGIYRGSNMTNWIIVPGSGNSGILLAQKFAEARNLGRITLQVLGEPWCVKINGISPNDY 379
Qy 452 TEAYLRVLQKVTADTQIVVCLLSSNEKDKYDAIKKYLCTDCPTPOCVVARTL-----505
Db 380 LEGKAAIKSVGDEIHHMLVVLADNKNTRYDSLKLYLCECPINQCVNLRITLAGSKD 439
Qy 506 -GKQOTVMAATKIALQMNCKMGGLWRVDIPLKVMIVGIDCYHDMTAGRRSIAGFVAS 564
Db 440 GGENKLSGIVLQMICTKGALWKNVPLKSTMIYGYDLYHDSLTKGTGACVST 499
Qy 565 INEGMTWFSRCIFQDRGOELVDGLKVCLOAALRAWNSCNE-YMPSRIIVYRDGVGDQL 623
Db 500 TSNDFTQFYSQTRPHENPTQLGNNLTHFVRKSLKQYYDNDNDKTLPSRLILYRDGAGDQI 559
Qy 624 KTLVNVYVPOFLDCLKSIGR-----GYNPLTVIVVKKVNRTRFFAQ--SGGRLQ 672
Db 560 PYIKNTEVLKVRDADCAVTDKAAELSNKVOEKIKLAFIIVTKVNMRIKQGSSEKSGAIN 619
Qy 673 PLPQTVIVDEVTREPWFVFFIVSQAVRSVSPHYNYVNSGLKPDHIOBLTYKLCHI 732
Db 620 PQGTVVDVTVTRPERMDFILVFPQVNGQVTVFVSYNIHDDTGLGDPDKHQQLAFKLC 679
Qy 733 YYNWPGVIRVPAPCOYAHKLAFLVQGSIHREPNLSNRLYYL 775
Db 680 YYNWQTVRVPAPCOYAHKLAFLTAQSLHDHANGYLRLDLPFL 722

RESULT 3
T41568
argonaute-like protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T41568
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.
submitted to the EMBL Data Library, May 1998
A:Reference number: Z21991
A:Accession: T41568
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-834 <WOO>
A:Cross-references: EMBL:AL023705; PIDN:CAAL19275.1; GSPDB:GN00066; SPDB:SPCC736.11
A:Experimental source: strain 972h-; cosmid c736
C:Genetics:
A:Map position: 3
A:Introns: 43/3
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 13.6%; Score 557.5; DB 2; Length 834;
Best Local Similarity 22.9%; Pred. No. 3e-32;
Matches 198; Conservative 155; Mismatches 346; Indels 167; Gaps 31;

Qy 20 GSSGIIVRLSTNHFRTSPQWALYOHIDYNPLMEARLRALLFOHEDL-----70
Db 16 GGLGKQITLKANFPQIISLNETINQYHVIVGDSRVPKRSQLIWNSEKVFYFGSSWM 75
Qy 71 -----IGKCHAFDGTILF-----LPKELQ--QKVTEV-----PSKTRNGED 104
Db 76 NSYDGRSMCWSEGDADGTIKVNIIGSEHPREIFSIQKSKINLHTLSQFVNSKISSD 135
Qy 105 VRITITLNLPTSPCTCLQFVNIIFRR-----LLKIMNQQIGRNYNENDPIDPSH 159
Db 136 -----PQVLSSTMFLLKLPSETLFGFMHSFTGNGVSLGGGVE-----178
Qy 160 LVTWPQFTTSILOYSIMLCTDVSHKVL-RSETVLDFM-----197

```


179 --AWKGFYQIRNQFMSVNVNDISSAFWRNDSLLQILMEYTDGCVNRDLTRFDLUKRUS 236

198 --FNFVHQTBEEHFKQEQVSKELIGLVLTKNXNKTYRVDIDWDQPKSTF-KKADGSE- 253

237 RKRFLKVTQCHRN-----VGTDLANRVYSIEGFSKSDASDFFVRLNGEQ 285

254 -VSFELEYRKOYNQOEITDLKQPVLVISQPKRRRGGGTLPGPAMLIPELCVLT---GLTOK 309

286 KISVAEYFLENHNVRQYFNLPCIL-----VKNGAMLPTEFCFVVKQORYTAK 333

310 MRNDENVMKDLAVHTRLTPQORQREYGRLLDYTHKND-NVQRELRDGLSFDNSNLISPSG 368

334 LNSD-----OTANKIRPAVORPERVOQIDDFVHQMDWDTPYLTOYGMIQKKMLEVPA 388

369 RILQTEKIHOGGKTDFYNPOFADWSKETRGAPLISV--KPLDNMLLI---YTRR-NYEA 422

389 RVLETPSIRYGGDIE-RPVSGRW-NLRGRFLDPPRAPIRSWAVMCFSTARLEPMRG 445

423 NSLIQNLFKVTPAMGQ--MEKAIMIEVDDR--TEAYLRVLQOKV---TADTQIVVCLL 474

446 ENFLQTVQTLTSLGFINFMKPKPVLVADIRGSVEELCITLYKKAQOVGNAPDYUFLFIL 505

475 SSNRKDKYDAIKKYLCTDCTPQCVVARTLGQOVTMAITKIALQMNCKMG--ELW 531

506 DKNSPEPYGSIKVCNTMLGVPSOCALSKHI--LQSKPOYCANLGMKINVKVGINCSLI 563

532 RVDIPLKLV--MIVGIDCYH-DMTAGRRSTAGFVASINEGHTWFSRCITFODRGOLVDG 588

564 PKSNPLGNVFTLJLGGDVHPGCVATGVSIAVASVDLNGCKYTVASRSQPHQVIEG 623

589 LKVCLOAALRAWNSCNEYMPSRIIVRGDVGDGKLTLVNVEYVQPFLDCLIKSGRYNPR 648

624 MKDIVVLLOGFRAMTKQQQRILYFDDGTSEGOFLSVINDELSQKEACHSLSPKYNPK 683

649 LTVIVVKRVNTRFFAGS---GGRLOMPLGTWIDVEVTRPEWYDFPIVSQAVRSGSVSP 705

684 ILVCTTQKRHHARFFIKNSGDGRNGNPLFTTIEKHVTHPEYOYDFFLISHPSLQGVSP 743

706 THYNVLYDNSGLKPDHILQRLTYKLCHYYNWPVGIVRVPACQVAHKLAFI----- 755

744 VHTVLHDEIQMPDQFQTLCYNLCTYARATSAVSLPVPYVYAHVLSNLRARYQDVTADD 803

756 -----VGQSTHREPNLSNRL 772

804 TFEVTSASMDQEV--KPLLALASKI, 827

RESULT 4

H86438
protein T19E23.7 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86438
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chan, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712

[illegible]

RESULT 5
T32079
hypothetical protein T07D3.7 - *Caenorhabditis elegans*
C/Species: *Caenorhabditis elegans*
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 21-Jan-2000
C/Accession: T32079
R/Lamar, B.; Wamsley, P.; Twyman, B.
submitted to the EMBL Data Library, July 1997

A:Description: The sequence of C. elegans cosmid T07D3.

A:Reference number: Z21121

A:Accession: T32079

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-891 <LAM>

A:Cross-references: EMBL:AF016682; PIDN:AA66187.1; GSPDB:GN00020; CESP:T07D3.7

A:Experimental source: strain Bristol N2; clone T07D3

C:Genetics:

A:Gene: CESP:T07D3.7

A:Map position: 2

A:Introns: 34/2; 92/2; 206/1; 560/1; 879/3

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 12.3%; Score 505.5; DB 2; Length 891;

Best Local Similarity 23.7%; Pred. No. 2e-28;

Matches 201; Conservative 139; Mismatches 338; Indels 169; Gaps 32;

QY 20 GSGGIIVRLSTNHRLTSRPMALYQYHIDYNPLMEARLRALLFQHEDLIGKCHAFDG 79

DB 52 GVEGRSILLRANHFVRI-PGGSVQHYQIDVPPDKCPRVNREVI-----G 96

QY 80 TILFLPKRLQOKVTEFSTRN-----GEDVRIITITNE-----LPTS 119

DB 97 CLISFSKYFTNIRPVYDGRNMYTRPLIGTEPMNFEVTLPGDSVERKFSVTMKWIG 156

QY 120 PTCLE-----QFYNIIFRLLKIMLQOIGRNYNPNPNDIPIDPSHRL 160

DB 157 QVCLSDALDAMEGRVQVPEHVAQSIDVLRH-LPSLKYPVGRSFFTPPGVMK-PGMQM 214

QY 161 -----VIWPGTTSILOVENSIMLCTDVSHKVL-RSETVLDFMNFVH----- 202

DB 215 HQBSKLGGRVWFGEFGHQSVPSCQWMLNIDVSATAFYRAMFVIEFAVLELPLVQALA 274

QY 203 -----QTEHKEQEQVSKELIGLVLTXY-----NNKTYRVDIDWDQPKSF-----KKA 249

DB 275 ERRLSDAQVKP-----THEIRGLKIEITHCGAVRRKIRVGNVTRRPAQTQTFPLQLETG 330

QY 250 DGSEVSEFLEYRKOYQNEITDLKOPVL-----VSQPKRRRGGPGTLPGPAMLIPELC----- 301

DB 331 QTEICTVAKYFFDKYR-----IQLVKPHPLCLOVGEQKH-----TYLPEVCDIVP 377

QY 302 ---YLGLTDKNDNFNMKDLAVHRLTPEQOREVGRGLDIYHKNDVNQREL----- 352

DB 378 GORCLKLTDVQSTW-----IKATARSAPR-RERBICKLVS-----KALSADPPA 423

QY 353 RDWGLSFDNLSFGRILOTEKIHOGG-KTPDYNPQFADWSKETRGAPLISVKPLDNW 410

DB 424 HEFGIITNPAMTEVKGRLVSAKLLYGGRRHRTALPNQGVW---DMRGKQFHTGMEVRTW 481

QY 411 -LLIYTRNVEAANSL---IONLFKVTTPANGMQ-----RKAIMIEVDDRTYAYLRVL 459

DB 482 AIACFAQOQSHVKENDLRFMTTQLOKISTDAGMPIIGTTPMFCYASGVE---QVEPMFKYL 538

QY 460 QQKVADTQVWCLSSNRKDKYDALIKKYLCTDCTPTSCQVARTLGGKQOTVMAATKIA 519

DB 539 KQYYSALQILVVVL---PGKTPVIAEVKRVGDTVLGIATQCVQAK---NATPTFQLSNLC 594

QY 520 LQNCNKGK-----ELWRVDIPLKLMIVGIDICYHDMTAGRR---SIAGFVASINEGMTR 571

DB 595 LKQNVKLGVSNIILPNVRPIEFNEPFIPLGCDITHPAAGDTRKPSIAAIVGSGMDAHPSR 654

QY 572 WFSRCIFQDRGOELVDGLKVCLOAALRAWNSCNEYMPSRILVYVDCVGGQOLKTNVVEV 631

DB 655 YATVRVQQRQELITDITMRELLVQFVTRTFKPARIVVTRDGVSGQFNVLQYEL 714

QY 632 PQFLDCLKSGIRGNPLTVIVKRVNTRFF-----AQSGRLQNPGLGTVDIVETRE 687

DB 715 RAIRKACVLESQVQPGITFIAVQKRHHTLPAADKADQVKAFFNPPTGTVGVTHPT 774

QY 688 WYDFEIVSQVRSGSVSPHNYIYNSGLKPHQRLITKLCHIIYNNWPGVIRVPAPCQ 747

DB 775 EEDPFLCSHAG1QGTSPSPSHYVLMDDNDITADELOOLTQOMCHTVYRCTRVSIPAPAY 834

QY 748 YAHKLAP 754

DB 835 YAHLVAF 841

RESULT 6

T22391

hypothetical protein F48F7.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000

C:Accession: T22391; T23234

R:Coles, L.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19559

A:Accession: T22391

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1000 <WIL>

A:Cross-references: EMBL:Z69661; PIDN:CAA93496.1; GSPDB:GN00028; CESP:F48F7.1

A:Experimental source: clone F48F7

R:Coles, L.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z19712

A:Accession: T23234

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1000 <WIL>

A:Cross-references: EMBL:Z69663; PIDN:CAA93512.1; GSPDB:GN00028; CESP:F48F7.1

A:Experimental source: Clone K02B9

C:Genetics:

A:Gene: CESP:F48F7.1

A:Map position: X

A:Introns: 70/3; 128/2; 185/2; 673/1

C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 12.2%; Score 498.5; DB 2; Length 1000;

Best Local Similarity 22.5%; Pred. No. 7.8e-28;

Matches 199; Conservative 139; Mismatches 322; Indels 225; Gaps 34;

QY 20 GSGGIIVRLSTNHRLTSRPMALYQYHIDYNPLMEARLRALLFQHEDLIGKCHAFDG 79

DB 145 GVEGRSILLRANHFVRI-PGCTIHYQVDTVP-----DKC----- 179

QY 80 TILFLPKRLQOKV-----TEVSKTRNGEDVRIITLTNELP-----PTSP-- 120

DB 180 -----PRVNVREIISCLISAFSKYFTNIRPVYDGRNMYTRPLIGRERMDPDLPGD 234

QY 121 -----TCLOFYNIIFRLLKIMLQOIGRN 145

DB 235 SAVERQPSVSKWVGVSLSLTLEDAMEGRVQVPPFAVQAMDVLRH-LPSLKYPVGRS 293

QY 146 YANPNDPIDIPSHRLV-----TWPGFTTSILQY 173

DB 294 FFSF--PVPNASGVWAGSCPPQASGAVAGSAHAGSAGVHAESKLGGRVWFGFHQSVRPS 351

QY 174 ENSIMLCTDVSHKVL-RSETVLDFMNFVH-----QTEHKEQEQVSKELIG 219

DB 352 QWRKMLNIDVSATAFYRAMFVIEFAVLELPLVQALAERALSADQKVPF-----TKERIG 407

QY 220 LVVLTKY---NNKTYRVDIDWDQNPKSTF---KKADGSEVSEFLEYRKOYQNEITDLK 272

DB 408 LKIEITHCGQMRKRYKRVNTRRPAQTQTFPLQLEGTQIECTVAKYFYDKYR---IQLK 464

QY 273 QPVL-----VSQPKRRRGGTLPGPAMLIPELCYLT-----GLTDKRNDFNVKDLA 321

DB 465 YPHLPCLQVQGEQKH-----TYLPEVCNIVPGQRCIKKLTDVQSTW-----IK 509

QY 322 VHTELTPEQOREVGRGLI-----DYIHKNDNVQKRLRDWGLSFDNLSFSGRILQI 373

DB 510 ATARSAPR-RERBISNLVRAKFAESADPPAF-----EFGITINPAMTEVKGVLSA 558

QY 374 EKHGGKKT-FDYNPQFADWSKETRGAPLISVKPLDNW-LLIYTRNVEAANSL---IQN 428

Db 559 PKLLYGRTRATAPALNQGWW--DMRGKQFHTGIDVRVWAIACFAQQOQHVKENDLRMETNQ 616
Qy 429 LFKVTPAMQMOM-----RKAIMLEVDRTAEYLRVLQOKVTADTQIVVCLLSNRKDK 481
Db 617 LQRIENDAGMPIVGNPCFKAVGVE--QVEFMPFKYLKONYSG-IQLVVVLV-PGKTPV 671
Qy 482 YDAIKKYLCTDPTSPQCQVARTLKGQOTVMAIAFKIALQNNCKMG--ELRWVDI 535
Db 672 YAEVXRVGDTVLGIATOCVQAK--NAIRTPQTLNCLKNVKGWVSGVILLNVPRI 729
Qy 536 PKLVMIVGIDYCHDMTAGRR--STAGFVASINEGRTWFRSCIPQDRGQBLVCLKVL 593
Db 730 FNEPVIFFGCDITHPPAGDSRPSIAAVGSDAHPYSIAATVRVQQRQBIISDLTVMV 789
Qy 594 QAAALRAWNSCNMPSRIIVYRDGVDGQKLTLYNNVYVQFELDLCKLSGRCVGNRLTIV 653
Db 790 RELVQFYNTREKPARIVVYRDGVSQGFNVLOYELRAIEACMMLRGVQGITFIA 849
Qy 654 VKKRVNTRFFA---QSGRLQNLPLPGTVIDVETRPBWYDFIVSQVRSVSPHTYN 709
Db 850 VQKRHHTRLFVAVDKQDVQKAYNIPPGTVDVGIHTPTTEFDYLCGHAGIQTSRPSHYH 909
Qy 710 VTYDNGSLKPDHQLRTLYKLYNNVPGVIRVAPCOVAHKLAF 754
Db 910 VLWDDNLTADLQQLTYQWCHTYVRCTRSVSIPAPAYAHVLVAF 954

RESULT 7
A84678
Argonate (AGOL)-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: A84678
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Mofat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: A84678
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-997 <STO>
A:Cross-references: GB:AE002093; NID:G4510428; PIDN:RAD21514.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g27880
A:Map position: 2
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 12.1%; Score 495.5; DB 2; Length 997;
Best Local Similarity 24.1%; Pred. No. 1.3e-27;
Matches 200; Conservative 139; Mismatches 363; Indels 129; Gaps 30;
Qy 20 GSSGIIVRLSTNHFRITSPQWALYQYHIDYNPLMEA---RLRSALLFOHED--LIGK 73
Db 154 GTLGGKVMVRANHF-LVQVADRLYHYDVSINPEVISTKNVNMVKKLVKNVKSGLGK 212
Qy 74 CHAFD-----GTILFLPKLQOKVTEVFSKTRNGEDVRITILTNELPTSTPCLQF 125
Db 213 SPAYDGRKSLYTAGLPFFDSKEFVNLAERKADGGSGKDRPFKAVKN---VTSTDLYLQ 269
Qy 126 YNIIIFRR-----LKKIMNL-----QOIGRNYNPNPDIPIPSHR-----LVI 162
Db 270 QQFLDRKQREAPYDTQVLVDVLRDPSNDYVSGRSPFHTSLGKADRGKELGDGITY 329
Qy 163 WPGFTTILQYNSIMLCTDVSHKVLRSETVL--DFMNFVHOTEHK-----FQGVSK 216
Db 330 WRGVFQSLRLTOMGLSLNIDVSARSFVEPIVTVDTFISKFLINRLNRLRDSRLKVKV 389
Qy 217 LIGL-VVLTKYN-NKTYRVDIDWDQNPSTFKKADGSEVSFLEYRYKQYNOEITLQOP 274
Db 390 LRLTKVLLHWNCTKSAKISGLSIFIRELURFTLEDKSEKTVVQYFAEKYINRVKQALP 449

Qy 275 VLVSOPKRRRGPGGTLPGPAMLIPELC-----YLTLGLTDQRNDFNMKDLAVHTRILT 327
Db 450 AIQT-----GSDTRPVYLPMLCOIDEQRYTKRLNEKO-----VTALLKATQOR 494
Qy 328 PQORQREVRLDLYHKNDNVORELDRWGLSPDNLSPSGLILQTE--KIHQGGKTPDY 385
Db 495 PPDRENSIKNLVVKNNYNDLSKE---FGMSTYTTQLASIEARVLPMLKYHDSGKRV 551
Qy 386 NPQFADWS---KETRGAPLISVKPLDNWLLI--VTRRNYEAANSLIQNLKFKVTPAMNQ 439
Db 552 NPLGQNMIDKKNVNGAKVTS-----WTCVSFSTRIDRGLPQEFCKOLIGMCVSKGWE 605
Qy 440 MRKA-----IMIEVDDRTAYLRVIOQKVATQIVVCLLSNRKDKYDAIKYLCDCP 494
Db 606 FKQPAIPFISCPPEHIEBALDI--HKRAPGLQLLIVIL-PDVTGSGYKIKRICETELG 662
Qy 495 TPOCVVARTLKGQOTVMAIAFKIALQNNCKMG--ELRWVDIPL--KLVMIVGI 545
Db 663 TVSQCCQPRQVKNLN--KOYMNVALKINVKYGRNTVNDAIRNIPITDRPTIIMGA 720
Qy 546 DCYHDMTA--GRRSIAGFVASIN--EGMTRWFRSCIPQDRGQBLVCLKVLQAALE--- 598
Db 721 DVTHPQPGSDSPSIAAVVASMDWPEINKYRGVLSAQAHREBIIQDLYKLVQDPQRLVH 780
Qy 599 -----ANNSCNEYMPSRIIVYRDGVDGQKLTLYNNVYVQFELDLCKLSIGRGNPRL 649
Db 781 SGLIREHFIATRRATGOIPORIIFYRDGVSGEQSVLLHEMTAIRKACNSIQENIVPRV 840
Qy 650 TVIWKRVNTRFFAQSOG-----RLQNLPLQGTVIDVEVTRPEWYDFIVSQVRSVSV 703
Db 841 TFVIVQKRHHTRLFPSOHGNRDNTDKSGNIQGTVDVTIKICHENEDFYLNHSHAGIQGS 900
Qy 704 SPHYNVIVDNGSLKPDHQLRTLYKLYNNVPGVIRVAPCOVAHKLAF 754
Db 901 RPAHYVHLLDENGFTADQLQMLTNLNCYTYARCTKSVSIVPPAYYAHLAAP 951

RESULT 8
T52134
Zwille protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C:Accession: T52134
R:Laux, T.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z25975
A:Accession: T52134
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-988 <LAU>
A:Cross-references: EMBL:AJ223508; PIDN:CAAL1429.1
C:Genetics:
A:Gene: zwille
A:Map position: 5
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 12.0%; Score 492; DB 2; Length 988;
Best Local Similarity 23.6%; Pred. No. 2.3e-27;
Matches 204; Conservative 133; Mismatches 397; Indels 132; Gaps 32;
Qy 5 VNTQNLDHVKESKTSGGIIVRLSTNHFRITSPQWALYQYHIDYNPLMEAARLSAL- 63
Db 120 MGVRKNSNFAPRFGFTGLTKCIVKANHF-LADLPKDLNQYDVTITPEVSSKVNRAII 178
Qy 64 -----LQCHEDLICKHAFD-----GTILFLPKLQOKVTE---VFSKTRNGEDVR 106
Db 179 AELVRLYKESDLGRPLPAYDGRKSLYTAGELPTFWKEFSVKIVDEDDGILNGPKRSRYK 238
Qy 107 ITIT-----LTNELPTSTPCLQYNIIFRLLKIMNLQOIGRNYNPNPDI 153
Db 239 VAIKVRANWHHLGFLAGKADRCQEAQVILDIVRB-LSVKRFPVGRSFPSPD--I 295

```

154 DIP---SHRLVWPQFTTSLIOYENSIMCLTDV-SHKVLRSETVLDWMFNFTQTEEHK- 208
296 KTPORLGBGLBSWCFCYOSIRPTOMGLSINTDMASAAFTIEPLVPVIEFVAQLLGKDVLSKP 355
209 ---FOEOVSKELLGLVLTXYN---NKTYVDDIDMDQNPSTFKKADGSEV-SFLEYI 260
356 LSDSDRVKIKGLRGKVETHRANVRKYRVAGLTTQPTRELMPFVDENCTMXSVIEYF 415
261 RKQYNQETIDLKQVPL-VSOPKRRRGPGGTIFGPAMLPIE-LCYLTGLTDMRDNFNMVK 318
416 QEMYGFTIQHTLPCLOVGNCK--ASYLPMERACKIVEGQRYTKRLNEQ----- 463
319 DLAVHTRLTPEORREVGRLTDYL----HKNDNVORELDRDWGLSPDSNLLSFGSRILQTE 374
464 ----ITALLKVTQRAQGRNDILRTVOHNAIDQDPYAKEFGMNISEKSLASVEARILPAP 519
375 --KIHOGGKTFDYNPQFADSKETRGAPLSVKVPLDNWLLIYTRRNYE--AANSLIQNLF 430
520 WLKYHNGEKEDCLPQVGQNMNMK--KMINGMTVSRWACVNPFSRVSQENVARGFCNELG 577
431 KVTPMQMQRKATMIEV----DRTE-----AYLRVLQQKVTDATQIVVCLLSNRKDK 481
578 OMCEVSGMETNPEBVIPIYSARPQVEKALKHVYHTSNKTKGKELELLAILPDNNGSL 637
482 YDAIKKYLCDCTPPSQCWAR---TLGKQOTVMAIATKIALOMNCKMGGE-----L 530
638 YGDLKRICEITELGLISOCCLTKHVFKSKQ-----YLADVSLKINVNNGGRTVLDAIS 692
531 WRV-----DIPKLWNVIGDICYHMTA--GRRSIAGFVASIN-EGMTRWFSCIFQDRGQ 583
693 CRPLVSDIP---TIIFGADVHPENGEESSPIAAVVASQDMPEVTKIAGLVCAQAHQ 749
584 ELVDGLKVCLQAALR-----ANWSCNEYMPSRITIVYRDGVGGOLKTLVNYE 630
750 ELIQDLKTYQDPRVGTVSGMIRDLLISFRKATGQKPLRIIFTRDGVSEGFQVQLIYE 809
631 VPQFLDCLKISIGRYNPRLTIVVYKRWNTFRFA-----QSGRLQNPLFGTVIDVEVT 684
810 LDAIRKACASLEPNYQPPVFIYVQKRHHTRLFANNHRDKNSTDRCNGLLPGTVVDTKIC 869
685 RPWYDFFIYSQAVRSQSVSPTHNNVIYDNSGLKPDHIORLTYKLCHYYNWPQVIRKPA 744
870 HPTEFPYLCSHAGIQGTSRAHHVLDWENNFTADGIQSLTNNLCYTACTRVSIVP 929
745 PCOYAHKLAPLVGQSIIHREPNLSLN 770
930 PAYAYHAAAPRA--RFYLEPEIMODN 953

```

RESULT 9
 Argonaute (AGO1)-like protein [imported] - Arabidopsis thaliana
 A84668
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C/Accession: A84668
 R./Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 403, 761-768, 1999
 A./Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A./Reference number: A84420; MUID:20083487; PMID:10617197
 A./Accession: A84668
 A./Status: preliminary
 A./Molecule type: DNA
 A./Residues: 1-930 <STO>
 A./Cross-references: GB:AE002093; NID:g3885334; PIDN:AAC77862.1; GSPDB:GN00139
 C./Genetics:
 A./Gene: At2g27040
 A./Map position: 2
 C./Superfamily: rabbit translation initiation factor eIF-2C
 Query Match 11.9%; Score 488.5; DB 2; Length 930;
 Best Local Similarity 21.3%; Pred. No. 3,7e-27;

	Matches	185;	Conservative	170;	Mismatches	340;	Indels	173;	Gaps	34;
QY	20	GSSGIIVRLSTNHPRLL	TSRPOWALYOYHI	----	DYNPLNEARRLBSALLFO	----	HE	68		
Db	63	GTRQKTPLLTNHFKVDVANLQGHFFHYSVALFVDDGRP	-VEQGVGKILDKVHOTYHS	121						
QY	69	DLIGKCHAFDG	-----TILFLPKRLQ	-QKVTEVFSKTR	-----NG	----	102			
Db	122	DLDCKEPAYDGEKILFTYIGALPSNMKDFSVLBEVSATS	KDFVSANGNSPNGNESPSD	181						
QY	103	-----EVRITITILNTELP	-----PTSPCTLOFYNIIPRLLKIM	137						
Db	182	GDKRLRRPNRSKFRVEISYAAKIPLOALANAMRGOESENQSEAIRVLDIILLQHAARQ	241							
QY	138	NLOQIGRNYNPNDDPIDP	-SHRLVIWPGFTTSLLOYENSIMLCTDV	-SHKVLRSFTVL	194					
Db	242	GCLLVROSFPFH	-NDPTNCEBPVGNILGCRGFHSSFRTTQGGMSLMDVTTMIIPKGPVV	300						
QY	195	DLMFNFYQTEEHKFP	-EQVSKELIGLIVLTKYNNKTYRVDDIDWDQNKSTFK	----	247					
Db	301	DFLIANQWARDPYSDWSKAKRTLKNLRVKVSPSGQEFKLTGLSDKPCROTTELKKNRP	360							
QY	248	-----KADGSEVSFLEYRKK	-----QYNQBITDLKQPLVLSQPKRRRGPGGTLPGPAML	297						
Db	361	NENGEFTTEVTADYPRDRHIDLOYSADLPCCIN	-----VGKPKR	----	406					
QY	298	PELC	-----YLTGTLTKMRNDFNMKDLAVHTRLTPEORQORQEVGGLDIYIHKNDNVQR	350						
Db	407	LELCALVPLORYTKALTFFORS	-----ALVKSROKPOERMVTLVSKALXV	-SNYDAEP	458					
QY	351	ELRDWGLSFDSNLLSPGRILQTEKIHQ	-GGKTFDYNPOFADKSKETRCAPLISVKPLD	408						
Db	459	LLRSCGISISSNTQVEGRVLPAPKLKMGCGSETFPRNGEWNNKE	-----FVEPTKIQ	513						
QY	409	NWLLIYTRRYEA	-----ANSLIQNLFKVTPAMQMOKRAIMIEVDD	----	455					
Db	514	RWVWV	-----NFSARCNNVRQVVDLLIKIGSGKGEIASPQVFEENQFRAPPIMKRVENN	569						
QY	456	LRVLQOKVITADTQIVCLLSSNR	-KDKYDAIKKYLCTDCTPSCQCVVARTLGHQQTVMAI	514						
Db	570	FKDIQSKLPQVPOFILCVLPDKNSDLYGFWKKNLTFFGIQVQCMAPTRQPNQDYL	----	626						
QY	515	ATKIALQMNCKMG	-----ELMRVD	-----IPLKLVMLVGIDCVHDMTAGRSIAGFV	562					
Db	627	-TWLLAKINAKLGLSNLSVETPAFTVISKVP	-----TILGMDVSHG	-SPGQSDVPESIA	681					
QY	563	ASINEG	-----MTKWFORCITPQDRQBELVQGL	-----KVCLOAALRAWNSCNYMPS	609					
Db	682	AVVSRREWPLISIKYRASVTRTQPSKAEMIESLVKNGTDEDDGIKELLVDFTYSSNKKRKE	741							
QY	610	RIIVRDGVDGOLKTLVANNVEPOFLDLKLSIGRGYNRLTVIVVKRVNTRFPAGSGR	659							
Db	742	HLIIFRDGVSQFNOVLNIEDQIIIEACKLDANNWPKPLLLVAQKNHHTKFFQPTSP	-800							
QY	670	LQNPFGTVIDVEVTRPEWYDFEIVSQVRSVSPTHYNNVIYDNSGLKPDHQRILTYKL	729							
Db	801	-ENVVPGTIIIDNKI	CHPKNNDYFLCAHAGMI	GTRTRTHYVLVDEYGFSADELOELVHSL	859					
QY	730	CHYYNWPQVIRVPACQVAHKLAFUNG	757							
Db	860	SYVYQRSTSAISVAPICVYAHAAALQ	887							

RESULT 10
JC6569 translation initiation factor eIF-2C - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C;Accession: PC5505; JC6569
R;Zou C.; Zhang, Z.; Wu, S.; Osterman, J.C.
Gene 211, 187-194, 1998
A;Title: Molecular cloning and characterization of a rabbit eIF2C protein.
A;Reference number: JC6569; MUID:98267198; PMID:9602122

A;Accession: PC6505

A;Molecule type: protein

A;Residues: 336-359;380-409;694-711 <ZOU1>

A;Experimental source: liver

A;Accession: J6569

A;Molecule type: mRNA

A;Residues: 1-813 <ZOU2>

A;Cross-references: GB:AF005355; NID:G3253158; PIDN:AA24323.1; PID:G3253159

A;Superfamily: rabbit translation initiation factor eIF-2C

C;Keywords: liver

Query Match 11.8%; Score 482; DB 2; Length 813;
Best Local Similarity 23.9%; Pred. No. 9e-27;
Matches 191; Conservative 147; Mismatches 341; Indels 120; Gaps 32;

```
QY 39 PMAIYQYHIDNPLMEARLSAL---LFOH--EDLTG-KCHADG-----TILFLPKR 87
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 PKIDYHVELDKPKCPRRVNRIVEHMHVQHFKAIFGDRKVPDGRKNLYTAMLEPIG 63
QY 88 LQOKVTEVFSKTRNGED--VRITI-----TLTNELPSTPTCLQFYNIERR 132
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 -REKVELEVTLFPEGKDRIFKYSIKWVSCVSLQALHDALSGRLPSVPETIQALDVNRH 122
QY 133 LKIMNLOQIIGNYINPDIDIP-SHRLVWPGETTSILOVENSIMLCTDVSHKVL-RS 190
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 -LPSMRKYTEVGRSFTTASGCGSNPLGGREVNFGHQSVRPSLWKWMLNIDVSATFYKA 181
QY 191 ETVLDF---MENFYHQTEHK-----FOQVSKELIGLVLTKY---NNKTYRVDIDWD 239
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 QPVIEFCEVLDKFKIECQKPLTDSQVVKFKTEIKGLKVEITHCGQMKRYKRVNTRR 241
QY 240 QNPKSTFKKADGS---EVSFLEYRQVNOBITDLKQVPL-VSQPKRRRGGTLPGPA 294
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 242 PASHQTFPLQESGQVETCTAQAQYFKDRHKLVLRYPLHPLCLQVQGEQKH-----T 291
QY 295 MLIPLCLVLTG-----LTDKQNDNFNMKDLAVHTLTPQORQREVRILDIYHKNDN 347
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 YLPLEVCNIVAGQRCIKLTD---NOTSTM--IRATARSAPD-RQEEISKLMR--SASFN 343
QY 348 VQRELRDGLSDSNLLSFSGRILQTEKHOGGKTFDYNQPADWSKETRGAPLSVKPL 407
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 344 TDPVYREFGIMVKDEMTDVTGRVLQPPSILYGR-----NKAIATPVQGVNDM 391
QY 408 DN-----WLLIYTRRYEAA-----NSLIQNLKFKVTPAMGMQMK-----AIMLEV 448
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 392 RNKQFHTGIEIKWAIACFAQRCQTEVHLKSFTEQLRKISRDAGNPQGGPCPKYAGG 451
QY 449 DRTEAYLRLVLOQKVTAQTQIVVCLLSNRKDKYDAIKKYLCTDQPTPSQCVAFTLGKQ 508
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 452 ADSVGPMPFRL-KNTYAGLQLVVIL-PGKTPVAEVRKRGDTVLGMATQCVQMKNV--Q 507
QY 509 QTVMAIAIKIALQWCKMGG-----ELWRVDIPKLVMYIGIDYHDMTAG---RRSIA 559
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 508 RITPQTLNCLCKINVLKGVNNILLPGRPFPVFOQPIVFGADVTHP-PAGDGKKPSIA 566
QY 560 GFVASINEGMRWFSRCIFQDRGOBLVDGLKYCLOALRAWNSCNEYMPSRIIVRDGVG 619
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 567 AVVGSMDAHPNRYCATVRVQQRHEIIQDLAMVRELLIOFYKSTRFPTRIIIFRDGVS 626
QY 620 DGQKLTLYNVEVPOFLDCLKSIGRGNPRLTVIVVKGKRWNTFFPA-----QSGRLQNEPLP 675
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 627 EGQFOQVLHLLAIREACIKLEKDYQPGITFIVVQKRHHTLFTCDKNERVSKSGNIPA 686
QY 676 GTVIDVEYTRPEWYDFVISOAVRSGSVSPHYNYINDNSGLKPDHIOQLTYKLCHIYNN 735
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 697 GTTVDTKLHTPEFDYFVLCSHAGIQGTSRPSGHYHVLWDDNRFSSDELQILTYQLCHTYVR 746
QY 736 WPGVLRVPAPQYAKHLAF 754
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 747 CTRSIVSIPAPAYAHLVAF 765
```

RESULT 11

T01113

translation initiation factor eIF-2C homolog T21L14.12 - Arabidopsis thaliana
N;Alternate names: Argonaute (AGO1)-like protein
C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 16-Feb-2001

C;Accession: T01113; D84739

R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kai
submitted to the EMBL Data Library, December 1997
A;Description: Arabidopsis thaliana chromosome II BAC T21L14 genomic sequence.

A;Reference number: Z14209

A;Accession: T01113

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-887 <ROU>

A;Cross-references: EMBL:AC003033; NID:G2702261; PID:G2702284

A;Experimental source: cultivar Columbia

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon, J.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: D84739

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-887 <STO>

A;Cross-references: GB:AE002093; NID:G2702284; PIDN:AAB91987.1; GSPDB:GN00139

C;Genetics:

A;Gene: At2g32940; T21L14.12

A;Map position: 2

A;Introns: 61/3; 130/2; 203/2; 248/1; 281/3; 314/2; 367/3; 406/3; 442/3; 463/3; 501/3;

C;Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.6%; Score 434; DB 2; Length 887;

Best Local Similarity 22.2%; Pred. No. 3.2e-23;

Matches 198; Conservative 153; Mismatches 349; Indels 190; Gaps 41;

QY 3 PGNVTRQNLDHVKESKVGSSGIIVRLSTNHFPLTSR-POWALYQYH-----DYNPLMEAR 57

D 1 :

Db 24 YDITTRGV-----GTTGNPIELCTNHFNSVRQPDVVFYQYTVTSITTENGADVDTG 75

QY 58 RLRSALLFO-----HEDLIGKCHAFDG-TILFLPKRLOOK-----VTEVFSKTRNG-- 102

D 1 :

Db 76 GISKLMQDLFKTYSSDLGDKRLAYDGEKTLTYVGPLPQNEFDELIVEGFSKRDGCVS 135

QY 103 -----EDVRITITITNELP-----PTSTCLQFYNI 129

D 1 :

Db 136 DGGSSSGTCKRSKRSFLPSRYKQIHVAAEFLKTLVLTQKRGAYTPDKSAQDALRLVDIV 195

QY 130 FRLLKTMNLOQIGRN-----YINPDPIIDPSHRLVWPGFTTSILOYENSIMLC 180

D 1 :

Db 196 LR-----QAAERGLLVRQAFHSDGHPMKVGGVIGI-RGLHSSFRPTHGGLSLN 246

QY 181 TDVS-HKVLRSSETVLDPMFNFYHOTEEHKFO---EQVSKELIGLVLTKYNNKTYRVD 235

D 1 :

Db 247 IDVSTTMLFPGPIEFL--KANQSVETPRQIDWKVAOKMLKMRVKAHNRNMEFKI-- 302

QY 236 IDWDONPKS---TFKKADGS-----EVSFLEYRQYNOEITDILKQVPL-VSQPKRRR 284

D 1 :

Db 303 IGLSSKPCNQOLFMSKIKDGEREVPRIETVYDFYKQTYTEPISSAYFPCLDVGKPDPR-- 360

QY 285 GPGGTLPGPAMLLPELCYLGLTDXKNDNFNMKDLAVHTP--LTPQORQREVR----- 337

D 1 :

Db 361 -----PNYLPFCNLSLQ-----RYTKPLSGRQVLLIVESSRQPKLERIKTLN 405

QY 338 --LIDYTHKNDNVQRELRDMLGSLFSDSNLLSFSGRILQTEKHOGGKTFDYNQPADWSKE 395

D 1 :

Db 406 DAMHTYCTDKDPF---LAGCGISLEKEMTQVEGRVLKPPML-KFGKNEDQPCNGRNFN 461

QY 396 TRGAPLISVKPLDNWLLIYTRRYE---AANSLIQNLKFKVTPAMGMOM-RKATMIEVD-- 449

D 1 :

Db 462 NK-MLEPRAIKSWAIV----NFSFPCDSSHISRELISCGMRKGIEIDRPFALVEEDPQ 515

[illegible]


```
QY 575 RCIFDRGOELVDGLKVCLOALRAWNSCNEYMPRIIVYRDGVGGQGLKTLVNYEVPQF 634
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
694 NVKQKCKRESVYLLDAIRERIIIFYRHTKQKPARIIIVYRDGVGGQFSEVLRREIQSI 753
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
635 LDCLKSIIRGYNPRLTIVVKKRVNTRPFAQ-----SGRLQNPPLGCTVIDVEVTRP 690
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
754 RTACLAIAEDFRPPITIVVQKRRHARIFCKYQNDMVGKAKVPPGTTVDGIVSPGEGF 813
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
691 PFIVSQAVRSVSPSTHYNVYDNSGLKPDHIQRLTYKLCHIIYNWPGVIRVPAPCOYAH 750
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
814 FYLCHYGVQGTSPARYHVLLDECKFTADEIQSIYGNCHYGRCTRSVSIPTPVYAD 873
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
751 KLAFIVGQSIHREPNSLSN 770
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
874 LVATRARCHVKRKLGLADNN 893

RESULT 15
D88568
protein ZK757.3 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: D88568
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; PMID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: D88568
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1040 <STO>
A:Cross-references: GB:chr_III; PIDN:CAA82941.1; PID:g3877004; GSPDB:GN00021; CESP:ZK757
C:Genetics:
A:Map position: 3
C:Superfamily: rabbit translation initiation factor eIF-2C

Query Match 10.3%; Score 421; DB 2; Length 1040;
Best Local Similarity 21.6%; Pred No. 3.6e-22;
Matches 199; Conservative 148; Mismatches 367; Indels 206; Gaps 36;

QY 20 GSGGIIVRLSTNHRL-TSRPQWALYQXHD-YNFLMEARRL---RSALLF-----QHED 69
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
93 GTIGRQIPVKSFFAMDLKXPKWVVIQXHVETHP--GCKLKDDBMRIFFWKAVSDHPN 150
QY 70 LIGK--CHAFDGT-ILFLPKRLQ-----OKVTEVSKTRNGEDVRITI----- 109
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
151 IFHNKFALAYDGAHQLYTVVARLEFPDDQGSVELDCEASLPKDNDRDTRCAISIQNVGPVL 210
QY 110 -----TLNLELPSTPTCLQFYNIIFRLLKIMNLQOIGRNYN-----PNDPI 153
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
211 LEMQRTINLNDERVLTIQILDIICROSLTCLPKXN-SANEYTWKSSCYRIPTAAGQAL 269
QY 154 DIPSHRLVWPQFTTSILOYEN-SIMCLTDVSHKVLRS--TVLDPNFNFYHQ----- 203
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
270 DLEGGK-EMWTGFFSSAHIASNYRPLLNDIVAHATFYKTRITVLQFMCDVLNERTSKPNR 328
QY 204 -----PCLHVG-----PTNIFLPMEHCLIDSPQYKXKXSEK 497
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
329 NNPRGPGGPGGGRGGRGSGSYGNFNGRPPGANVRDDFGNGLTFMTDLRSRDT 388
QY 205 EEHKEQEQVSKELI-GLVVLTKYNNKTVRVDIDWDQNP--KSTFKKAD--GSEV--SFL 257
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
389 QLSSEFTRIFGAIKGMKIRAAHRPNAIFVYKNSLIQPADKLMFQGDIEGRQVVCVA 448
QY 258 EYRKQYNQEIITDLKQPLVLSQPKRRRPGGTLPGPAMLIP-ELC-----YLTGLTDK 309
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
449 DYFSEKYG---PLKVPKL---PCLHVG-----PTNIFLPMEHCLIDSPQYKXKXSEK 497
QY 310 MENDFNWVKDLAVHRLTFPEQRQREVGRLLDIYHKNDNVQRELRDNLGSLFSDNLLSFSGR 369
```

```
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
498 QTS--AIKKAADA-----TOREDRIKOLA--AQASFCTDPFLKEFGVAVSSQMIQTAR 549
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
370 ILQTEKIHOGKTFDYN-----POFADWSKETRGAPLSVKPLDNWLLIYTRNTEAANSL 425
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
550 VIQPPPIFMFGNNRSVNPVFPKDGSWTMDNQTLYMPATCRSYSMIALVDRDQTSLOTF 609
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
426 IQNLFKVTTPAMGQMRK-----AIMIEVDDRTAEAYLRVLOOKVTAQT-QI 469
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
610 CQSLTKRATAMGNFPWPDLVKYGRSKEDVCTLTFLIAD-----EYRVINTVDCD 660
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
470 VVCLLSNRKDKYDAIKYLCTDCTPTSOQVARTLKGQQTVMALATKIALOMNCKMG- 528
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
661 IIVVLQSKNSDIYMTVKEQSDIVHGINSQCVMKNVSR--PTPATCANIVLKLANKMGGI 718
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
529 -----ELWRVDIPLKLVMIVGIDCYHDMTAGRR-----SIAGFVASINEGTRWFS 574
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
719 NSRIVADKITNKILVDQB--TWVVGIDVTHPTQAEHRMMWMPVAAIVANVDLLPQSYGA 775
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
575 RCIFDRGOELVDGLKVCLOALRAWNSCNEYMPRIIVYRDGVGGQGLKTLVNYEVPQF 634
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
776 NVKQKCKRESVYLLDAIRERIIIFYRHTKQKPARIIIVYRDGVGGQFSEVLRREIQSI 835
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
635 LDCLKSIIRGYNPRLTIVVKKRVNTRPFAQ-----SGRLQNPPLGCTVIDVEVTRP 690
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
836 RTACLAIAEDFRPPITIVVQKRRHARIFCKYQNDMVGKAKVPPGTTVDGIVSPGEGF 895
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
691 PFIVSQAVRSVSPSTHYNVYDNSGLKPDHIQRLTYKLCHIIYNWPGVIRVPAPCOYAH 750
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
896 FYLCHYGVQGTSPARYHVLLDECKFTADEIQSIYGNCHYGRCTRSVSIPTPVYAD 955
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
751 KLAFIVGQSIHREPNSLSN 770
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
956 LVATRARCHVKRKLGLADNN 975
```

Search completed: December 4, 2003, 19:16:09
Job time : 26 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 19:08:48 ; Search time 18 Seconds

(without alignments)
2024.761 Million cell updates/sec

Title: US-10-043-774B-2

Perfect score: 4102

Sequence: 1 MIFGNTFRQLDHVKESKGTG.....VQSIHREPNSLSNRLYYL 775

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	557.5	13.6	834	1	IF2C SCHPO
2	510	12.4	1048	1	AG01 ARATH
3	506.5	12.3	857	1	I2C1 HUMAN
4	504.5	12.3	863	1	I2C2 RAT
5	503	12.3	988	1	PINE ARATH
6	495.5	12.1	987	1	AG01 ARATH
7	482	11.8	813	1	I2C2 RABIT
8	421	10.3	1040	1	I2C3 HUMAN
9	374	9.1	377	1	YO53 CAEEL
10	248.5	6.1	1032	1	DICE HUMAN
11	137.5	3.4	1912	1	DICE MOUSE
12	134.5	3.3	1906	1	DICE MOUSE
13	118	2.9	1120	1	MFD RICPR
14	114	2.8	2035	1	HMP2 YEREN
15	113.5	2.8	560	1	PHAC PSOL
16	113	2.8	949	1	PMAC ARATH
17	113	2.8	3144	1	VP13 YEAST
18	108	2.6	1317	1	CLH YEAST
19	108	2.6	1653	1	CLH YEAST
20	106.5	2.6	690	1	Y173 URBPA
21	104	2.5	1180	1	C4AA BACTI
22	104	2.5	1277	1	PDS5 YEAST
23	104	2.5	4128	1	FRD5 YEAST
24	103	2.5	1343	1	RPOB HAIN
25	103	2.5	1575	1	IQG2 HUMAN
26	102.5	2.5	354	1	CHLI CHLVU
27	102.5	2.5	869	1	YD95 YEAST
28	102.5	2.5	1170	1	ITR2 BOVIN
29	102	2.5	3122	1	DPO2 MOUSE
30	101.5	2.5	3135	1	S230 PLAF0
31	101	2.5	1388	1	RPOD TOBAC
32	101	2.5	1886	1	POL COYMY
33	100.5	2.5	1162	1	VGL2_IBVE

RESULT 1
ID IF2C SCHPO STANDARD; PRT; 834 AA.
AC 074957;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable eukaryotic translation initiation factor 2C (eIF2C) (eIF-2C).
GN SPCC736.11.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S., Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Mungall K., Murphy L., Niblett D., Ostell C., Oliver K., O'Neill S., Pearson D., Quay M.A., Rabinowitsch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Skelton J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Voiclaert G., Aert R., Robben J., Grymonprez B., Welljens I., Vanstreels B., Rieger M., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H., Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas M., Rochet M., Gallard C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L., Carrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J., Sipakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RL "The genome sequence of Schizosaccharomyces pombe."
RT Nature 415:871-880(2002).
CC -I- FUNCTION: Plays an important role in the eukaryotic peptide chain initiation process (by similarity).
CC -I- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -I- SIMILARITY: Contains 1 PZ domain.
CC -I- SIMILARITY: Contains 1 Pfam domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

34 100.5 2.5 1169 1 EX5B BORBU
35 100 2.4 963 1 KINH HUMAN
36 100 2.4 1014 1 NEBL HUMAN
37 100 2.4 1587 1 TOP2_PENCH
38 99.5 2.4 879 1 MISP_DROME
39 99.5 2.4 1358 1 SIR4 YEAST
40 99.5 2.4 4568 1 DYHC CAEEL
41 99 2.4 749 1 STA4 MOUSE
42 99 2.4 806 1 SYL_BACHD
43 99 2.4 944 1 DNL4 YEAST
44 99 2.4 1086 1 SYL RICPR
45 99 2.4 2670 1 IP3T RAT

ALIGNMENTS

```
CC or send an email to license@sib.ch).
CC -----
DR EMBL; AL023705; CAA19275.1; -.
DR PIR; T41568; T41568.
DR GenDB SPombe; SPC736.11; -.
DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; P1WI.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; P1WI; 1.
DR PROSITE; PS08221; PAZ; 1.
DR PROSITE; PS08222; P1WI; 1.
KW Initiation factor; Protein biosynthesis.
FT DOMAIN 212 326 PAZ.
FT DOMAIN 500 799 P1WI.
SQ SEQUENCE 834 AA; 94438 MW; F5C397BF590D9E4 CRC64;

Query Match
Best Local Similarity 13.6%; Score 557.5; DB 1; Length 834;
Matches 198; Conservative 155; Mismatches 346; Indels 167; Gaps 31;

QY 20 GSGIIVRLSTNHRFTSPQWALYQYHIDYNFLMEARRLSALLFOHEDL----- 70
DQ 16 GGLGKQITLKANFFQIISLPNETINQYHIVGDSVPRKQSLIWNSEKVFQFGSSW 75
QY 71 -----IGKCHAFDGTILF-----LPKELQ---QKVTEV-----FSKTRNGED 104
DQ 76 NSVYDGRSMCWSGDADGTIKVNGISESHPREIEFSIQKSKINILHTLSQFVNSKYSD 135
QY 105 VRITITLNLPTSPSTCLOFYNIIFRR-----LLKIMNLQIQGRNKNPNDPIDIPSHR 159
DQ 136 -----PQVLSIMFLDLILKKRSETLFGFMSEFFTGENGVSLLGGVE----- 178
QY 160 LVTPGFTSILQYENSIMLCTDVSHKVL-RSTVLDFM----- 197
DQ 179 --AWKGFYSIRENCGFMSVNVDISSAFWNDSLLQILMEYDCGNVRDLTRFDLKLRS 236
QY 198 --PNFVHQTSEHKFQOVSKELGLVLTQYNNKTYRVDIDWDQPKSTF-KKADGSE- 253
DQ 237 RKRFELKVTQCHRN-----VGTDLANRVYSIEGFSKSDSDFVRLRNGEEQ 285
QY 254 -VSFLEYRYKQYNOETDLKQPLVSPQRRRPGGTLPGPAMLIPELCVLT---GLTDK 309
DQ 286 KISVARYFLENHVRVQYENLPCL-----VKGAMLPFEFVVGQRYTK 333
QY 310 MRNDFNMKDLAVHTLTPEQREYGRGLDYTHKND-NVQRELRLDWGLSFDNLSLFSG 368
DQ 334 LNSD-----QTANMIRFAVQRPPEVQQIDDFVHQMDWDTPYLTQYGMKIQKMLEVEA 388
QY 369 RILQTEKIHQGGKTFDYNQFADMSKETRCAPLISV--KPLDNWLLI---YTR-NYEA 422
DQ 389 RVLETSPISRYGGDIE-RPVSGRWN--LRGRFLDPPRAPIRSWAVWCFTSTRRLPMRGI 445
QY 423 NSLIQNLKVTYPAMGQ--MRKAIMIEVDNR--TEAYLRVLOQV---TADTQIVVCLL 474
DQ 446 ENFLQYVQVLTSLGFINFMVKPPVLYADIRGSEBELCITLYKKAQGVNAPDYLFFIL 505
QY 475 SSNRKDYDAIKYLCTDCTPSCQVAVPLGKQQTMAITKIALQMNCKMG---ELW 531
DQ 506 DKNSPEFYSGIKVCNCTMLGVPSCQAIKSHI--LQSKPOYCANLGMKINKVGGVINCSLI 563
QY 532 RVDPILKLV--MIVGIDCVH-DMTAGRRSTAGFVASINEGWTWFSRCIFQDQGLVDG 588
DQ 564 PKNPLGNVFTLIGGVDPHPGATGVSTASIVASVDLNGCKYTAVSRSQPRHQVEIEG 623
QY 589 LKVLQALRAWNSCNEYMPSRIIVRGDVGDLTKTLVNYEYVQFQDLCLKSTGRGNPR 648
DQ 624 MKDIVVLLQGFAMTKQQRILYIFRDGTSEGOFLSVINDELISOKEACHSLSPKYNK 683
QY 649 LTVIVVKRVNTRPFAQS---GGRLONLPLGTVIDVEVTPPEWDFPIVSVQAVRSVSP 705
DQ 684 ILVCTTQKRHRHARFFIKNKGSDGRNGNPLPGTITIEKHTVHPYQYDFVILSHPSLQGVSP 743
QY 706 THYNYVDNSGLKPDHRTQRTYKLCIHYNNPVGIVPAPCOYAKHLAF----- 755
```

```
Db 744 VHYTVLDEIQMPDQFQTLNCLYVYATRSVSLVPEVYAHVLSNLARYQDVTADD 803
QY 756 -----VGQSIHREPNLSLNRL 772
DQ 804 TFEVTESEASMDQEV--KPLALSLSSKL 827

RESULT 2
AGOI ARATH
ID AGOI ARATH STANDARD; PRT; 1048 AA.
AC O04379.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Argonaute protein.
GN AGOI OR ATIG48410 OR FLIA17.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
ON NCBI TaxID=3702;
RX [1]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Leaf;
RX MEDLINE=98090460; PubMed=9427751;
RA Bohmert K., Camus I., Bellini C., Bouchez D., Caboche M., Benning C.;
RT "AGOI defines a novel locus of Arabidopsis controlling leaf
development.";
RL EMBO J. 17:170-180(1998).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Alcafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Krenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Millscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana.";
RL Nature 408:816-820(2000).
CC 1- FUNCTION: ESSENTIAL FOR PROPER DEVELOPMENT OF LEAVES AND FLORAL
CC ORGANS AND FORMATION OF AXILLARY MERISTEMS.
CC 1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC 1- TISSUE SPECIFICITY: WIDELY EXPRESSED AT LOW LEVELS.
CC 1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT ALL DEVELOPMENTAL
CC STAGES.
CC 1- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC 1- SIMILARITY: Contains 1 PAZ domain.
CC 1- SIMILARITY: Contains 1 P1WI domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; U91995; AAC18440.1; -.
DR EMBL; AC007932; AAD49755.1; -.
```

DR InterPro; IPR003100; PAZ.
DR InterPro; IPR003165; Piwi.
DR Pfam; PF02170; PAZ; 1.
DR Pfam; PF02171; Piwi; 1.
DR PROSITE; PS0821; PAZ; 1.
DR PROSITE; PS0822; PIWI; 1.
KW Developmental protein.
FT DOMAIN 391 501 PAZ.
FT DOMAIN 676 997 PIWI.
FT DOMAIN 13 104 GY-RICH.
SQ SEQUENCE 1048 AA; 116190 MW; 385146343A09C541 CRC64;

Query Match 12.4%; Score 510; DB 1; Length 1048;
Best Local Similarity 25.1%; Pred. No. 4.8e-29;
Matches 213; Conservative 136; Mismatches 376; Indels 124; Gaps 33;

QY 20 GSSGIIVRLSTNHFRITSPQWALYQYHIDYNPLMEARLRSLALPQHEDLIGKCH--- 75
Db 186 GQSGKRCIVKANHF-FAELPKDLHHYDVTITPEVTSRGVNRVAVMKQLVDNRYRDSLHLSR 244

QY 76 --AFDG-----TILFLP-----KRLQOKVTEFSKTRNGEDVRIIT 109
Db 245 LPAYDGKSLYTAGLPFFNSKEPRINLLDEEVGAGQRREREPKVIKLVARADLHLGM 304

QY 110 TLNLEPPTSPCTCLQFYNIIFRLLKIMNLQOIGRNYNPNPIDIPSHR-----LVIWP 164
Db 305 FLEGKSDAPQALQVLDIVLR-LFTSRYPVGRSFYSP-----DIKKQSLGDGLSNR 359

QY 165 GFTTSILOYSIMLCTDVSHKV-LRSETVLDPMFNFYHQTSEHK-----FOQVSKELI 218
Db 360 GFYOSIRPTQMLSLNIDMSSTAFIEANPVIQVCDLLNRDISRPLSDADRVIKKALR 419

QY 219 GL-VLITKYN--KTRVDDIDWDQPKSTFKAD-GSEVSFLEYKQYNQBITLQKP 274
Db 420 GVKVEVTHRGNNRKYRISGLTAVATRELTPPVDERTOKSVVEYHETVGFRIQHTQLP 479

QY 275 VLVSQPKRRRGPGCTLPGPAMLPCLCYLT--GLTDQRNDNMKDLAVHTRLTPEQRQ 332
Db 480 CLOVGNLSR-----PNYLPMEVCKIVGQRYSKELNERQITALLKV-TQRPIDRE 529

QY 333 REV---GRLLDYTHKNDNVORELDMGLSFDLSNLLSFSGRILOTE--KIHQGGKTFDYNP 387
Db 530 KDILQTVQLNDY--AKDNYAQE---FGIKISTSLASVEARILPPPMKLVHESGREGTCLP 584

QY 388 OFADWSKETRGAPLISVKPLDMLLIYTRNYP--AANSLIQLNFKVTPANGQMR-KAI 444
Db 585 QVQGMNMNKK--KHINGGTNNMICINFSRQVQDNLTARTFCBELAQCYVSGMAFNPEPV 642

QY 445 MIEVDRTTEAYLVLOQKVTADT-----QIVVCLLSNRKKDYDAIKKYLCTDCPT 495
Db 643 LPVVSARPEQEVKVLKTRYHDATSKLSQGEIDLLIIVLPDNGSLYGLDKRICETELGI 702

QY 496 PSQCWARTLGK-QQTWAIATKIALQMKMG-ELMRVD-----IPL---KLVMIVGI 545
Db 703 VSQCCLTKHFKASKQYMA---NVALKINVKVGRNTVLVDALSRRPLVSDRPTIFGA 759

QY 546 DCVHDMTA--GRSISAGFVASIN-BGTRWFSRCIFQDRGOELVDGI----- 589
Db 760 DVTHPHGEDSSPSIAAVASQDMPEITTKYAGLVCAQAHQELQDLFKENWKDPQKGVVT 819

QY 590 KVCLOALRAWNSCNEYMPRIIVYRGVGDGLKTLVNTVEVPOFLDCLKSIGRYNPLR 649
Db 820 GMIKELLIAFRSTGHKPLRIIFYRGVSEGQYQVLLVELDAIRKACASLEAGYQFPV 879

QY 650 TVIIVKRVNTRFFAQ-----SGGRQLNPLPGTVIDVETRPWEYDFFIVSOVRSVSV 703
Db 880 TFVVQKRRHTRLFAQNHNDHSVDRSGNLPGLTVVDSKICHPTEPDFYLCSHAGIQTGS 939

QY 704 SPTHNVINYNSGLKPDHIOQLTKLCHYYNWPGEVTRVAPCOYAKHLAFLVQSQSHRE 763
Db 940 RPAHYHVLWDENNFTADGLQSLTNLCTVARTCRSVISVPPAYYAHAAFA--RFYME 997

QY 764 PNLSLSNRL 772

Db 998 PETSDSGSM 1006

RESULT 3
ID I2C1 HUMAN STANDARD; PRT; 857 AA.
AC Q9ULI8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Eukaryotic translation initiation factor 2C 1 (eIF2C 1) (eIF-2C 1)
DE (Putative RNA-binding protein Q99).
GN EIF2C1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20005943; PubMed=10534406;
RA Roosters R., Adams V., Betts D., Moos R., Schmid M., Siermann A.,
RA Haastem S., Weitz S., Lichter P., Heitz P.U., von Knebel Doeberitz M.,
RA Briner J.;
RT "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
genomic organization, localization to chromosomal bands 1p34-p35, and
expression.";
RL Genomics 61:210-218(1999).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
initiation process.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AF093097; AAF00068.1; --
CC Genew; HGNC:3262; EIF2C1.
CC MIM; 606228; --
CC GO; GO:0003850; C:eukaryotic translation initiation factor 2...; TAS.
CC GO; GO:0003743; F:translation initiation factor activity; TAS.
CC DR; GO:0006446; P:regulation of translational initiation; TAS.
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003165; Piwi.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC PROSITE; PS0821; PAZ; 1.
CC PROSITE; PS0822; PIWI; 1.
KW Initiation factor; Protein biosynthesis.
FT DOMAIN 226 346 PAZ.
FT DOMAIN 515 816 PIWI.
SQ SEQUENCE 857 AA; 97214 MW; 1DBE524AE7CBAP66 CRC64;

Query Match 12.3%; Score 506.5; DB 1; Length 857;
Best Local Similarity 24.1%; Pred. No. 6.5e-29;
Matches 203; Conservative 160; Mismatches 347; Indels 133; Gaps 37;

QY 20 GSSGIIVRLSTNHFRITSPQWALYQYHIDYNPLMEARLRSLALPQHEDLIG-K 73
Db 30 GTVGKPIKLLANYFEV-DIPKIDVHYEVDIKPKPRVNRREVVEYVQHFQKPFQDGR 88

QY 74 CHAFDG-----TILFLP---KRLQOKVTEFSKTRNGED--VRITI-----T 110
Db 89 KPVIDGKKNYITVATLPIGNERVDFEVT---IPGEGKDIRFKVSIKWLAIVSWRLHEA 144

QY 111 LTNELPPTSPCTCLQFYNIIFRLLKIMNLQOIGRNYNPNPIDIP-SHRLVWPQFTTS 169

DB 145 LVSGQIPVPLESVQALVDVNRHILAS-MRYTPVGRSFFSPPEGYHPLDGGREWFVGFHOS 203
QY 170 ILQYENSIMLCTDVSHKVL-RSETVDFMP-----NFVHQTEBEHFKFORV--SKELIGL 220
DB 204 VRPAMKMLNIDVSATAFYKAQPVIEFCEVLDIENIDEPKPLTDSQVRFTKEIGL 263
QY 221 VVLTKY---NNKTYRVDIDMDGNPKSTF---KKADGSEVSFLEYRQYQNOEITDLKQ 273
DB 264 KVEVTHCGGMKRYRVCNTRRPASHQTFPLQESQGVTECTVAQYFKQKYNLQ---LKY 320
QY 274 PVL-----VSQPKRRRGGTLPAPMLIPELCVLTG-----LTDKMRDNFNMKDLAV 322
DB 321 PHLPCLOVGOEQKH-----TYLPLEVCNIVAGQICIKLTD---NOTSTM--IKA 365
QY 323 HTRLTPEQRQREVRGLIDYIHKNDNVQRELDRWGLSFDGNLISFSGRIILQTEKHGGKT 382
DB 366 TARSAPD-RQEISRLMK--NASYNLDPIQEFGIKVKDMTEVTGRVLPAPILQYGGN 422
QY 383 FDY-NPQFADWSKETGAPLISVKPLDNMLLIYTRNRYEANSLION----LFKVTPAMG 437
DB 423 RAIAATPNQGV--DMRGKQFYNGIEIKVWAIACFAPQKQCREVLKNFTDLRKISKDAG 480
QY 438 MOWRK---ALMIEVDDRTYALRVLOQKVTADTQIVVCLLSSNRKDYDAIKKYLCTDC 493
DB 481 MPTQGOFCFKYAGADSVEPNFRHLKNTYSG-LQIIVILV-PGKTPVAEVRKVGDTLL 538
QY 494 PTPSQCVARTLGKQOTVMAIATKIALQMNCKMGBELMRVDIP-----LKLWMIVGID 546
DB 539 GWATQCVQKVVVK--TSPQTLNGLKINVKLGG-INNLVPHORSAYFQPVIFLGAD 595
QY 547 CYHDMTAG---RRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAWNSC 603
DB 596 VTHP-PAGDGKPSITAVVSGMDAHPSCYCATVRVORPROEITIEDLSYVRELLIOFYKS 654
QY 604 NEYMPSRITVYRGDGGQGLKTAVNVEVQFOLDCLXSIGRGNPRLTIVVKKRVNTRFE 663
DB 655 TRFKPRIIFYRGVVEGQPLQIHVELLAIACIKLEKQYQPGITYIVVQGRHRLRF 714
QY 664 A-----QSGRLQNLPGTVIDVETRPPEWYDFIVSOAVRSVSPSTHNYIYDNSGLXP 719
DB 715 CADKNERIGKSGNIPAGTVDVNTWTHPEFDFYLCSHAGIQGTSRPSHYVVLWDDNRFTA 774
QY 720 DHIQRLTYKLCHLYNNWGVIRVPACQYAHKAF-----LVGOSIH 761
DB 775 DELQILTYQLCHTYVTRCTRSVSIAPAYARLVAFRARYHLVDKEDSGSGSHISGOSG 834
QY 762 REP 764
DB 835 RDP 837

RESULT 4
ID 12C2 RAT STANDARD; PRT; 863 AA.
AC Q9QZ81;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2)
DE (Goldi ER protein 95 kDa) (GERp95).
GN EIF2C2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Hepatoma;
RX MEDLINE=99443791; PubMed=10512872;
RA Cikaliuk D.E., Tabbaz N., Hendricks L.C., DiMattia G.E., Hansen D.,
RA Pilgrim D., Hobman T.C.;
RT "GERp95, a membrane-associated protein that belongs to a family of
RT proteins involved in stem cell differentiation."

RL Mol. Biol. Cell 10:3357-3372(1999).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL, AF195534; AAF12800.1; -.
CC InterPro: IPR003100; PAZ.
CC InterPro: IPR003165; Piwi.
CC Pfam: PF02170; PAZ; 1.
CC Pfam: PF02171; Piwi; 1.
CC PROSITE: PS0821; PAZ; 1.
CC PROSITE: PS0822; PIWI; 1.
KW Initiation factor; Protein biosynthesis.
FT DOMAIN 239 352
FT DOMAIN 521 822
FT SEQUENCE 863 AA; 97659 MW; 9A1E0D6B84F7B622 CRC64;
QY 20 GSSGIIVRLSTNHPRLTSRPOWALYQYHIDYNPLMEARLSAL---LPQH--EDLIG-K 73
DB 36 GTTGRTIKLQANFFEM-DIPKIDIVHYELDIKPEKPRVRNREIVEHVMVQHEKTFQIGDR 94
QY 74 CHAFQG-----TILFLPKLOOKYVEVFSKTNGED--VRITI-----TLTN 113
DB 95 KPVFGRKNLYTAMFLPIG-RDKLEVLTPGEGKDRIFKVKISKWVSCVSLQALHDALSQ 153
QY 114 ELPPTSPTCLQFVNIIFRLLKIMLQOIGRNYNPNPIDIP-SHRLVWPGFTTSILQ 172
DB 154 RLPSVFFETIQALDVVMRH-LFSMRYTPVGRSFFPTASEGCSNPLGGREWFVGFHOSV 212
QY 173 YENSIMLCTDVSHKVL-RSETVDF---MFNFYHOTEHK-----FQEQVSKELIGL 223
DB 213 SLWKMLNIDVSATAFYKAQPVIEFCEVLDIENIDEPKPLTDSQVRFTKEIGL 272
QY 224 TKY---NNKTYRVDIDMDGNPKSTFKKADGS-----EVSFLEYRQYQNOEITDLKQ 276
DB 273 ITHCGOMKRYRVCNTRRPASHQTFPLQESQGVTECTVAQYFKDRHKLVLRYPHLPCL 332
QY 277 -VSQPKRRRGGTLPAPMLIPELCVLTG-----LTDKMRDNFNMKDLAVHRLTP 328
DB 333 QVGOEQKH-----TYLPLEVCNIVAGQICIKLTD---NOTSTM--IRATARSAP 377
QY 329 EQRQREVGRLIDYIHKNDNVQRELDRWGLSFDGNLISFSGRIILQTEKHGGKTDFY 388
DB 378 D-RQEISKLNR--SASFNTDFYREFGIMVWDEMVDVTRGVLPQPSILYGR----- 427
QY 389 PADWSKETGAPLISVKPLDN-----WLLIYTERNYEAA-----NSLIQNLFKYT 433
DB 428 -----NKAITAPVQGVWDMRNKQFHITGIEIKVWAIACFAPQKQCREVLKNFTDLRK 482
QY 434 PAMGQMRK-----AIMIEVDDRTYALRVLOQKVTADTQIVVCLLSSNRKDYDAIKKYL 489
DB 483 RDAGMPIQGPCFCYKAGADSVEPNFRHL-KNTYAGLQVLVVIL-PGKTPVAEVRKVG 540
QY 490 CTDCPTPSCVARTLGKQOTVMAIATKIALQMNCKMGB-----ELMKRVDIPLKLVMTIV 543
DB 541 DTVLGMATQCVQKVVVK--TSPQTLNGLKINVKLGG-INNLVPHORSAYFQPVIFLGAD 598
QY 544 GIDCDVHDMTAG---RRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAW 600

Query Match 12.3%; Score 504.5; DB 1; Length 863;
Best Local Similarity 24.1%; Pred. No. 9.2e-29;
Matches 197; Conservative 151; Mismatches 349; Indels 121; Gaps 33;
QY 20 GSSGIIVRLSTNHPRLTSRPOWALYQYHIDYNPLMEARLSAL---LPQH--EDLIG-K 73
DB 36 GTTGRTIKLQANFFEM-DIPKIDIVHYELDIKPEKPRVRNREIVEHVMVQHEKTFQIGDR 94
QY 74 CHAFQG-----TILFLPKLOOKYVEVFSKTNGED--VRITI-----TLTN 113
DB 95 KPVFGRKNLYTAMFLPIG-RDKLEVLTPGEGKDRIFKVKISKWVSCVSLQALHDALSQ 153
QY 114 ELPPTSPTCLQFVNIIFRLLKIMLQOIGRNYNPNPIDIP-SHRLVWPGFTTSILQ 172
DB 154 RLPSVFFETIQALDVVMRH-LFSMRYTPVGRSFFPTASEGCSNPLGGREWFVGFHOSV 212
QY 173 YENSIMLCTDVSHKVL-RSETVDF---MFNFYHOTEHK-----FQEQVSKELIGL 223
DB 213 SLWKMLNIDVSATAFYKAQPVIEFCEVLDIENIDEPKPLTDSQVRFTKEIGL 272
QY 224 TKY---NNKTYRVDIDMDGNPKSTFKKADGS-----EVSFLEYRQYQNOEITDLKQ 276
DB 273 ITHCGOMKRYRVCNTRRPASHQTFPLQESQGVTECTVAQYFKDRHKLVLRYPHLPCL 332
QY 277 -VSQPKRRRGGTLPAPMLIPELCVLTG-----LTDKMRDNFNMKDLAVHRLTP 328
DB 333 QVGOEQKH-----TYLPLEVCNIVAGQICIKLTD---NOTSTM--IRATARSAP 377
QY 329 EQRQREVGRLIDYIHKNDNVQRELDRWGLSFDGNLISFSGRIILQTEKHGGKTDFY 388
DB 378 D-RQEISKLNR--SASFNTDFYREFGIMVWDEMVDVTRGVLPQPSILYGR----- 427
QY 389 PADWSKETGAPLISVKPLDN-----WLLIYTERNYEAA-----NSLIQNLFKYT 433
DB 428 -----NKAITAPVQGVWDMRNKQFHITGIEIKVWAIACFAPQKQCREVLKNFTDLRK 482
QY 434 PAMGQMRK-----AIMIEVDDRTYALRVLOQKVTADTQIVVCLLSSNRKDYDAIKKYL 489
DB 483 RDAGMPIQGPCFCYKAGADSVEPNFRHL-KNTYAGLQVLVVIL-PGKTPVAEVRKVG 540
QY 490 CTDCPTPSCVARTLGKQOTVMAIATKIALQMNCKMGB-----ELMKRVDIPLKLVMTIV 543
DB 541 DTVLGMATQCVQKVVVK--TSPQTLNGLKINVKLGG-INNLVPHORSAYFQPVIFLGAD 598
QY 544 GIDCDVHDMTAG---RRSIAGFVASINEGTRWFSRCIFQDRGOELVDGLKVCLOALRAW 600

```
Ds 599 GADVTHP-PAGDGKKPSIAAAGVSGMDAHPNRYCATVRVQQHROEIIQDLAAVRELLIQF 657
Qy 601 NSCENYMPRIIIVYRCVGGQKLTIVNVEVDFLCLXSGRGYNPRLTVIWKXVNT 660
Db 658 YKSTRFKPTRIIFYRDGVSEGQFQVHLHELLAIREACIKLEKXQPGITFIVQKRHT 717
Qy 661 RFFA---OSGRLQNPCTVIDVTRPEWDFPIVSOAVRSQVSPTHYVINDSG 716
Db 718 RLFCCTKXNVRGKSGNIPAGTIVDTKITHTEPFDYLCSHAGTQTSRSHVHLWDDNR 777
Qy 717 LKPDHQRITLYKLCHYYNMPGVIRPAPCOYAKLAF 754
Db 778 FSSDEQLITLYQLCHYYVRCRSVSPAPAYYALVAF 815

RESULT 5
PINH ARATH
AC Q9XGW1; O49256; STANDARD; PRT; 988 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE PINHEAD protein (ZWILLE protein).
GN AT5G43810 OR MQD19.17.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euroside II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Landsberg erecta;
RX MEDLINE=99094873; PubMed=9876176;
RA Lynn K., Fernandez A., Aida M., Sedbrook J., Tasaka M., Masson P.,
RA Barton M.K.;
RT "The PINHEAD/ZWILLE gene acts pleiotropically in Arabidopsis
RT development and has overlapping functions with the ARGONAUTE1 gene.";
RL Development 126:469-481(1999).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Landsberg erecta;
RX MEDLINE=98169396; PubMed=9501101;
RA Moussian B., Schoof H., Raeker A., Juergens G., Laux T.;
RT "Role of the ZWILLE gene in the regulation of central shoot meristem
RT cell fate during Arabidopsis embryogenesis.";
RL EMBO J. 17:1799-1809(1998).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DEVELOPMENTAL PROTEIN REQUIRED FOR RELIABLE FORMATION OF
CC PRIMARY AND AXILLARY SHOOT APICAL MERISTEMS. MAY BE A COMPONENT OF
CC A HYPOTHETICAL MERISTEM FORMING COMPETENCE FACTOR.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 PZ domain.
CC -!- SIMILARITY: Contains 1 PZ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC ENBL; AF154272; AAD40098.1; -
CC ENBL; AJ223508; CAAL1429.1; -
CC ENBL; AB026651; BAB11310.1; -
CC PIR; T52134; T52134.
CC InterPro; IPR003100; PZ.
```

```
DR InterPro; IPR003165; P1wi.
DR Pfam; PF02170; PZ; 1.
DR Pfam; PF02171; P1wi; 1.
DR PROSITE; PS50821; PZ; 1.
DR PROSITE; PS50822; P1wi; 1.
KW Developmental protein.
FT DOMAIN 337 451
FT DOMAIN 625 946
FT CONFLICT 475 479
FT CONFLICT 671 671
FT CONFLICT N -> D (IN REF. 2).
SQ SEQUENCE 988 AA; 110867 MW; 32EBB349C613DA20 CRC64;

Query Match 12.3%; Score 503; DB 1; Length 988;
Best Local Similarity 23.7%; Pred. No. 1,4e-28;
Matches 204; Conservative 136; Mismatches 399; Indels 122; Gaps 32;

Qy 5 VNTRQNLHVKESKTGSSGIIIVRLSTNHFRLTSRFPQWALYQYHIDYNPLMEARRLSAL- 63
Db 120 MGVKNSNFAPRPGFGTLTKCIVKANHF-LADLPTKDLNQYDVTITPEVSSKSVNRAII 178
Qy 64 -----LQHEDLIGKCHAFD-----GTILFLKRLQOKVTE-----VFSKTRNGEDVR 106
Db 179 AELVRLYKESDLGRRLPAYDGRKSLYTAGELPFTWKFSVKIVDEDDGLINGPKRERSYK 238
Qy 107 ITIT-----LTNELPPTSPTCLOFYNIIFRLLKIMNLQOICRNVNPDPI 153
Db 239 VALKFAVANMHLGEFLAGKADCPQAVQLDVLRE-LSVKFCPCVGRSFFSPD--I 295
Qy 154 DIP---SHRLVIWPGFTTSILOYENSIMLCTDV-SHKVLRSETVLDPMFNFYHQTEHK- 208
Db 296 KTFORLGELESWCIFYQSIRPTOMGLSLNIDMASAAPIELPLVFIEFAQLLGKDVLSKP 355
Qy 209 ----FQEQVSKELIGLVLTKN---NKTYRVDIDMDQNPSTKTKADGSEV-SLEYX 260
Db 356 LSDSDRVKIKKGLRGVKVEVTHRANVRKRYVAGLTTQTRHEIMFPVDENCTMKSVEYF 415
Qy 261 RKQYNQEIITDLKQVPL-VSQPKRRRPGGTLPGLPAMLIPELCYLGLTKDKNDFNMKD 319
Db 416 QEMYGFTQHTLPLCLOVGNQK---ASYLPWEACKIVEGQRYT----KRLNEKOITAL 467
Qy 320 LAVHTLTPEQOREVGRLLIDYIHKNDNVQRLDWGLSFDNSNLSFSGRIILQTE--KTH 377
Db 468 LKV-TCQPRDRENDILRTVQ--HNAYDQDPYAKEFGMNISEKLSAVEARILPAPWLKYH 524
Qy 378 QSGKTFDYNQPADMSKETRGAPLISVKPLDMWLLIYTERNYE--AANSLIQNLKFKVTA 435
Db 525 ENGKXKDCPLPQVGQNMNMNK--KMINGMTVSRWACVNFSSVQENVARGFCHELGQMEV 582
Qy 436 MGXQWREKAIMIEV---DDRTS-----AYLRVLQKQVTADTQIVVCLSSNRKDKYDAIK 486
Db 583 SCWENPEPVIPIYSARPDPVEKALKHVYHTSNKTKGKELLLAILPNNGLSLYGLDK 642
Qy 487 KYLCTDCPTSPQCVAR---TLGKQQTVAIATKIALQNCXNGE-----LMRV-- 533
Db 643 RICETELGLISOCCLTKHVFKISKQ-----YLANVSLKINVMGGRNTVLVDAISCIPL 697
Qy 534 --DIPKLVMYIGIDCYHDMTA--GRRSTAGFVASIN-EGMTWFSRCIFODRGQELVDG 588
Db 698 VSDIP---TIIFGADVTHPENGEESSPSIAAVASQDWPEVTYAGLVCAQAHROBLIQD 754
Qy 589 LKVLQQAALR-----ANSCNEMPSRIIVYRDGVGQGLTKLVNVEYVQPL 635
Db 755 LKVTQDPVRGTVSGMTRDILLISFRKATQKQPLAIFVRDGVSEGFQVLLYELDAIR 814
Qy 636 DCLKSIGRGYNPRLTVIWKXVNTFFFA-----QSGRLQNPFGTVIDVEVTRPEWY 689
Db 815 KACASLEPNYQPEVTFIVQKSHHTLFPANNHRDKNSTDRSGNILLPGTVVDTKICHPTEF 874
Qy 690 DFTFVSQAVRSQVSPTHYVINDSGLKPDPHQRITLYKLCHYYNMPGVIRPAPCOYA 749
Db 875 DFYLCSHAGTQTSRSHVHLWDDNFTADGIQSLTNLCTYARCTRSVSVIPPAYYA 934
Qy 750 HKLAFIVGQSIHREPNI.SLSN 770
```



```

Db      935 HLAAPRA--RFYLEPEIMQDN 953

RESULT 6
AGOL ARATH
ID AGOL ARATH STANDARD; PRT; 997 AA.
AC Q9SJK3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Argonaute-like protein A22927880.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsid.
OX NCBI_TaxID=3702;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Kechum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Unayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nickman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana.";
RL Nature 402:761-768 (1999).
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AC006929; AAD21514.1; -
CC PIR; A84678; A84678.
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003165; Piwi.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC PROSITE; P850821; PAZ; 1.
CC PROSITE; P850822; PIWI; 1.
CC FT DOMAIN 359 471 PAZ.
CC FT DOMAIN 638 958 PIWI.
CC SEQUENCE 997 AA; 111134 MW; DA93A1446C42F31 CRC64;

Query Match 12.1%; Score 495.5; DB 1; Length 997;
Best Local Similarity 24.1%; Pred. No. 5.1e-28;
Matches 200; Conservative 139; Mismatches 363; Indels 129; Gaps 30;

QY 20 GSGGIIVRLSTNHRFTSRPQWALYQYHIDYNPLMEA-----RRLRSAFLQHEH--LICK 73
Db      154 GTLGGKVMVRANHF-LVQVADRLYHDVYSINDEVISKTVRNRVMKLVKNYKDSHLGCK 212
QY 74 CHAFD-----GTLFLPKRLQKQVTEVFSKTRNGEDVRITITNELPTSPCLQF 125
Db      213 SPAYDGRKSLYTAGLPFFDSKEFVNVLAERKADGGSGKORPFKVAVN---VISTDLYQL 269
QY 126 YNIIFRR-----LLKINWL-----QQIGENYNPNDEIPSHR-----LVI 162
Db      270 QQFLDRQREAPYDTIQVLVDVLRDPSNDYVSGSFHTSLGKDARDGRGELGDIY 329

```

```

QY 163 WPGFTTSLQYENSIMLCTDVSHKVLRSSETVL-DFMNFYHQTEBHK-----FQGVSKSE 216
Db      330 WRGYFQSRLRLTQMGSLNIDVSARSFYEPFVVTDFISKFLNIRDNLRLPLRDSRLKVKV 389
QY 217 LIGL-VLTKYN-NKTYRVDDIDNDONPKSFKADGSEVSEFLEYRQYNOSEITDLKOP 274
Db      390 LRTLKVLKLLHWNCTKSAGISSIPRELFTLEDKSEKTVVQFAEKYRNVRYQALP 449
QY 275 VLVSQPKRRRGPGGTLPGFAMLIPELC-----YLTGLTKDRNDENVMDKLAVHRLT 327
Db      450 AIQT-----GSDTRPVVLPHELCOIDEGQRTYKELNEKQ-----VTALLKATCOR 494
QY 328 PEQOREVGRLLIDYIHKNDNVQRELRLDWGLSFDNSLLSFSGRILQTE--KIHGGKTFDY 385
Db      495 PPDRENSIKNLVKNVYNDLSKE---FGMSVTTQLASIEARVLPPPMKLVKHSCKERNV 551
QY 386 NPOFADWS---KETRGAPLISVKEPLDNWLLI--YTRRYEAAANSLLIQNLKFAVTAMNQ 439
Db      552 NPLRGQMNNDKQWNGAKVTS-----WTCVSFSTRIDRGLPQEFCKQLIGMCVSKGME 605
QY 440 MRKA-----IMIEVDDRTEAYLRVLQKVTADTQIVVCLLSNRKDKYDAIKKYLCTDCP 494
Db      606 FKQPAIFPISCPPEHIEALDI--HKRAGLQLLIVIL-PDVTGSGYKIKRICETELG 662
QY 495 TPSQCQVARTLGKQOTVMATKIALQNCXKGG-----ELWRVDIPL---KLWIVGI 545
Db      663 IVSQCCQPRQVKNL--KQYMNVAIKNVKTGGRNTVLNDAIRRNIPLITDRPTIIMGA 720
QY 546 DCYHDMTA--GRRSTAGFVASIN-EGMTFWPSCRIFDQGOELVDGLKVCLOAAR--- 598
Db      721 DVTHPQGEDSPSIAAVASMDWPEINKYRGLVSAQAHRHEIIOQLKVLVDOPORGLVH 780
QY 599 -----AMNSCNEYMPRIIVYRDGVGQGLKTLVNYEVPQFLDCLKSIGRGNPLR 649
Db      781 SGLIREHFIARRAQIPQRIIFRFDGVSEQPSQVLLHMTAARKACNSLOENYVPRV 840
QY 650 TVIVVKVNRTRFPAQSGG-----RLQNPILPGTVIDVEVTRPEWYDFIVSQAVRSGSV 703
Db      841 TFVIVQKRRHTLPEQHGNDMTKSGNIQGTVDVTKICHNEFEFVYLNHSHAGTQGS 900
QY 704 SPTHNVLYDNGSLKPDHIOBLTYKLCHLYNWPQVIRVPAPQYAHKLAF 754.
Db      901 RPAHYHLLDENGFTADQLQMLTNLCYARCTKSIVSIVPPAYVLAALAF 951

RESULT 7
I2C2 RABIT
ID I2C2 RABIT STANDARD; PRT; 813 AA.
AC O77503;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Eukaryotic translation initiation factor 2C 2 (eIF2C 2) (eIF-2C 2).
GN EIF2C2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1] SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Liver;
RX MEDLINE=98267198; PubMed=9602122;
RA Zou C., Zhang Z., Wu S., Osterman J.C.;
RT "Molecular cloning and characterization of a rabbit eIF2C protein.";
RL Gene 211:187-194 (1998).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC initiation process.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```


D_b 687 GTTVDTKITHTPEFDVLCSHAGIGGTRPSHYHLWDDNRPSSDELQILTYQLCHYYVR 746

Q_y 736 WFGVIRVPAPCQYAKLAF 754
 : : ||| ||| : |||

D_b 747 CTRSIPAPAYAHLVAF 765

```

RESULT 8
ID Y043 CAEEL STANDARD; PRT; 1040 AA.
AC P34681;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK757.3 in chromosome III.
GN ZK757.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodirinae; Caenorhabditis.
NCBI_TaxID=6239;
RN [1]_
RX
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RC MEDLINE=94150718; PubMed=7906398;
RX Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fuxton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightnig J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on it
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement
CC or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; Z30215; CAA82941.1; -
CC EMBL; Z29121; CAA82941.1; JOINED.
CC EMBL; Z29121; CAA82389.1; -
CC EMBL; Z30215; CAA82389.1; JOINED.
CC FIR; D88568; D88568.
CC WormPep; ZK757.3; CE01117.
CC InterPro; IPR003100; PAZ.
CC InterPro; IPR003165; Piwi.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC PROSITE; PS00821; PAZ; 1.
CC PROSITE; PS00822; PIWI; 1.
CC Hypothetical protein.
KW DOMAIN
FT DOMAIN 378 486
FT DOMAIN 660 966
FT PIWI.
SQ SEQUENCE 1040 AA; 115415 MW; 4C0483C1F1D72338 CRC64;

Query Match 10.3%; Score 421; DB 1; Length 1040;
Best Local Similarity 21.6%; Pred. No. 1.4e-22;
Matches 199; Conservative 148; Mismatches 367; Indels 206; Gaps 36

QY 20 GSGGIIVELSTNNHRL-TSPQWALXYQYHD-YNPLMEARL---RSALLF-----QHED 69
Db 93 GTIGRQIVKSNFFAMDKNPKMVIQYHYVEIHP--GCKLKDQMRVIFPKAVSDHPN 150

```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN SEQUENCE FROM N.A.
RP MEDLINE=20005943; PubMed=10534406;
RX Koesters R., Adams V., Betts D., Moos R., Schmid M., Siermann A.,
RA Hassam S., Weitz S., Lichter P., Heitz P.U., von Knebel Doeberitz M.,
RA Briner J.;
RT "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
RT genomic organization, localization to chromosomal bands 1p34-p35, and
RT expression.";
RL Genomics 61:210-218(1999).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC -!- INITIATION PROCESS.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 PwI domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF121255; AAF13034.2; -
CC Genew; HGNC:3263; EIP2C2.
DR GO; GO:0003743; P:translation initiation factor activity; NAS.
DR GO; GO:0006412; P:protein biosynthesis; NAS.
DR InterPro; IPR003165; PwI.
DR Pfam; PF02171; PwI; 1.
DR PROSITE; PS0822; PwI; 1.
KW Initiation factor; Protein biosynthesis.
FT NON_TER 1
FT DOMAIN 35 336 PwI.
FT SEQUENCE 377 AA; 42529 MW; 3BCE91689D58526 CRC64;
SQ
Query Match 9.1%; Score 374; DB 1; Length 377;
Best Local Similarity 30.2%; Pred. No. 8.5e-20;
Matches 96; Conservative 64; Mismatches 140; Indels 18; Gaps 7;
QY 450 DRTAYLRVLOQKVADTQIVVCLSSNRKDKDAIKYLCTDCPTPQCQVVARILGKQQ 509
Db 17 DSVEPMFRHL-KNTYAGLQVVL-PGKTPVYAEKRVGDTVLGNATQCVQMKNV--QR 72
QY 510 TWYATATKALQNMCKMG-----ELMRVDIPLKLVMTVGIDCYHDMTAG---RESIAG 560
Db 73 TTFQTLNMLKINVKLGQVNNILLPQGRPPVFBFVIFLGADVTHP-PAGDGKPKPSIAA 131
QY 561 FVASINEGTRWFSRCIFQDRGOELVDGLKVLQALRAWNSCNEYMPSPRIIVYRDGVD 620
Db 132 VVGRMDAHPNRYCATRVQQRHQEIQLDLAAMVRELLIQFYKSTRFKPRIIFYRDGVSE 191
QY 621 GQLKTLVNVDPQDLCKLSIGRGYNPRILTIVLVKRVNTRFFA-----QSGRLQNPLEG 676
Db 192 GQFQVLLHLELAIREACIKLEKDYQPGITFTVWQKRHHRTLFCTDKNERVKGSGNIPAG 251
QY 677 TVIDVEVTRPEWYDFIVSQAVRSQVSPHYNVYDNSGLKPDHIOQLTYKLCHYYNW 736
Db 252 TVTDYKITHPTFEFDYLCSHAGIQGTSRPSHYHLVNDNRFSDELQILTYQLCHYYVRC 311
QY 737 PGVIRVPAPCOYAHKIAF 754
Db 312 TRSVSIPAPAYVAHLVAF 329
RESULT 10
Y053 CAEEL
ID Y053 CAEEL STANDARD; PRT; 1032 AA.
AC Q09249;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN SEQUENCE FROM N.A.
RP MEDLINE=20005943; PubMed=10534406;
RX Koesters R., Adams V., Betts D., Moos R., Schmid M., Siermann A.,
RA Hassam S., Weitz S., Lichter P., Heitz P.U., von Knebel Doeberitz M.,
RA Briner J.;
RT "Human eukaryotic initiation factor EIF2C1 gene: cDNA sequence,
RT genomic organization, localization to chromosomal bands 1p34-p35, and
RT expression.";
RL Genomics 61:210-218(1999).
CC -!- FUNCTION: Plays an important role in the eukaryotic peptide chain
CC -!- INITIATION PROCESS.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 PwI domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF121255; AAF13034.2; -
CC Genew; HGNC:3263; EIP2C2.
DR GO; GO:0003743; P:translation initiation factor activity; NAS.
DR GO; GO:0006412; P:protein biosynthesis; NAS.
DR InterPro; IPR003165; PwI.
DR Pfam; PF02171; PwI; 1.
DR PROSITE; PS0822; PwI; 1.
KW Initiation factor; Protein biosynthesis.
FT NON_TER 1
FT DOMAIN 35 336 PwI.
FT SEQUENCE 377 AA; 42529 MW; 3BCE91689D58526 CRC64;
SQ
Query Match 9.1%; Score 374; DB 1; Length 377;
Best Local Similarity 30.2%; Pred. No. 8.5e-20;
Matches 96; Conservative 64; Mismatches 140; Indels 18; Gaps 7;
QY 450 DRTAYLRVLOQKVADTQIVVCLSSNRKDKDAIKYLCTDCPTPQCQVVARILGKQQ 509
Db 17 DSVEPMFRHL-KNTYAGLQVVL-PGKTPVYAEKRVGDTVLGNATQCVQMKNV--QR 72
QY 510 TWYATATKALQNMCKMG-----ELMRVDIPLKLVMTVGIDCYHDMTAG---RESIAG 560
Db 73 TTFQTLNMLKINVKLGQVNNILLPQGRPPVFBFVIFLGADVTHP-PAGDGKPKPSIAA 131
QY 561 FVASINEGTRWFSRCIFQDRGOELVDGLKVLQALRAWNSCNEYMPSPRIIVYRDGVD 620
Db 132 VVGRMDAHPNRYCATRVQQRHQEIQLDLAAMVRELLIQFYKSTRFKPRIIFYRDGVSE 191
QY 621 GQLKTLVNVDPQDLCKLSIGRGYNPRILTIVLVKRVNTRFFA-----QSGRLQNPLEG 676
Db 192 GQFQVLLHLELAIREACIKLEKDYQPGITFTVWQKRHHRTLFCTDKNERVKGSGNIPAG 251
QY 677 TVIDVEVTRPEWYDFIVSQAVRSQVSPHYNVYDNSGLKPDHIOQLTYKLCHYYNW 736
Db 252 TVTDYKITHPTFEFDYLCSHAGIQGTSRPSHYHLVNDNRFSDELQILTYQLCHYYVRC 311
QY 737 PGVIRVPAPCOYAHKIAF 754
Db 312 TRSVSIPAPAYVAHLVAF 329
RESULT 10
Y053 CAEEL
ID Y053 CAEEL STANDARD; PRT; 1032 AA.
AC Q09249;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
```

DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C16C10.3 in chromosome III.
DN C16C10.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RA Lloyd C.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: STRONG, TO C.ELEGANS C14B1.7. ALSO TO C06A1.4,
CC F49F7.1, R06C7.1, T22B3.2 AND ZK757.3.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 246787; CAA86741.1; --
CC PIR; T19324; T19324.
CC WormPep; C16C10.3; C01494.
CC InterPro; IPR003100; PAZ.
CC Pfam; PF02170; PAZ; 1.
CC Pfam; PF02171; Piwi; 1.
CC PROSITE; PS50821; PAZ; 1.
CC PROSITE; PS50822; PIWI; 1.
KW Hypothetical protein.
FT DOMAIN 369 481
FT DOMAIN 650 977
FT PIWI.
SQ SEQUENCE 1032 AA; 117105 MW; EBEAE23CE67BB149 CRC64;

Query Match 6.1%; Score 248.5; DB 1; Length 1032;
Best Local Similarity 20.9%; Pred. No. 4.9e-10;
Matches 131; Conservative 98; Mismatches 229; Indels 169; Gaps 24;

QY 206 EHKFOEQ-----VSKELIGL-VLTKYNNKTVYDDIDWDQPKSTFKKADGS---EVS 255
DB 381 EHAFGERTKCEIEALKGLDVECTHLKGNLRVSSIAENNAENTSFMKDKDGERVT 440
QY 256 FLEYTRKQVNEITDLKQVIVSQPKR-----RKGPGTLPGPAMLIPELCYLTG 305
DB 441 VAEYFLLQVNIKLKYPRLPLVWSKPKHESFFPMELLRIAPQRIKVNKNSPTVQSAMTG 500
QY 306 LTDKMRDNFNMKDLAVHTLTPQORQREVGLDIYHKNDNVQRELDRWGLSFPDSNLLS 365
DB 501 RNASH-----POHVKLVQDIL-----RDNKLEQNKYMDAPFGIKLMS 538
QY 366 -----FSGRILQTEKHGGKTF--DYN--PQADWSKETRGAPL--ISVKPLDNLWLI-- 413
DB 539 TEPIQWAKLLPQAKFKGTQYMPDMSEPAFTQDKFVEPARIRKIGIVFDNCIQMRQ 598
QY 414 -----YTRNYEAANSL-IQNLFKVPAMPQMRKAIMIE 447
DB 599 AEDFCDKLSNFCRDNGITVEKDSRDWSIRELNSDSVAIONLMKKW----- 644
QY 448 VDDRTEAYLRVLOQKVATQIVVCLLSNRKDKYDAIKY-----LCTDPTPS 497
DB 645 LDDRVDILVGIAREK-----KPDVHDILKYFESIGLQIQLC----- 682
QY 498 QCVWARTLKGQ---QTVMAITKIALQNMCKMGELWRVDIP----- 536
DB 683 QQTVDKMGCGGGRQTIIDNMRK-----FNLKCGGNFFVEIENAVRGKAVCSNNTLRKK 738
QY 537 -LKLVMIVGIDCYHDMTA-----GRSIAAGFVASINEGMTRWFSRCIFQDRGOE 584

DB 739 LLEHVQFIFGFEISHGASRTLFDPRSQMDGSPSVVGVSVSLTNS-TQLGGFTYLQTOKEY 797
QY 585 LVDGLKVCLOALRAWNSCNEMPSRIIVYRDGVDGQGLKLVNVEVQFLDCLKSIGRG 644
DB 798 KLOKLDEFFPCVRSYKSHSKTLPTRIVIVRYGAGEGNFNR-VKEEVEEMRTTFDKIQPG 856
QY 645 YNPLRTIIVWKKRVNTRFFAQ--SGRL--QNPLPGTVIDVEVTRPEWVDRPIVQAVRS 700
DB 857 YRPHLVIIAQASHARVFPSCISGNRATQNIPTGTCVENVLTSYGDFELSSQTPLI 916
QY 701 GSVSPHYNVIVDMSGLKPDHIORLTY 727
DB 917 GTVRPKYITLVNDKAKSKNELMHTY 943

RESULT 11
DICE HUMAN
ID DICE HUMAN STANDARD; PRT; 1912 AA.
AC Q9UPY3; O95943; Q9U002;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoribonuclease Dicer (EC 3.1.26.-) (Helicase with RNase motif)
DE (Helicase-MOI).
GN DICER1 OR DICER OR HERNA OR KIAA0928.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20246304; PubMed=10786632;
RA Matsuda S., Ichigotani Y., Okuda T., Irimura T., Nakatsugawa S.,
RA Hamaguchi M.;
RT "Molecular cloning and characterization of a novel human gene (HERNA)
RT which encodes a putative RNA-helicase.";
RL Biochim. Biophys. Acta 1490:163-169(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Provost P., Dishart D., Doucet D., Hermansson A., Frendewey D.,
RA Samuelsson B., Radmark O.;
RT "RNA binding and processing by recombinant human Dicer.";
RN Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
RN [4]
RP REVISIONS.
RX MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual
RT curation of 330 KIAA cDNA clones.";
RL DNA Res. 9:99-106(2002).
RN [5]
RP SEQUENCE OF 1238-1912 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=99162526; PubMed=10051563;
RA Provost P., Samuelsson B., Radmark O.;
RT "Interaction of 5-lipoxygenase with cellular proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:1881-1885(1999).
CC -!- FUNCTION: Involved in cleaving double-stranded RNA in the RNA
CC interference (RNAi) pathway. It produces 21 to 23 bp dsRNAs
CC (siRNAs) which target the selective destruction of homologous
CC RNAs.
CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY.

QY	210	Q8VQSKELIGL	-----VVLTKYNN-----KTYRVDDIDWDQNPFS	244
Db	893	---SEARIGIPSTKYTKETPFVKLEDYQDAVILPIRYRNPDPQHRFYVADVYTLDTPLS	948	
QY	245	TFKKADGSEVSFLEYRYKQYQNEITDLQPVU	-----VSQ	279
Db	949	KFPSPSEYB--TFAEYTKYKYNLDLNLNQPLLDVDHTSSRLNLLTPRHNLNQKGPALPLSS	1006	
QY	280	PKRRGPGGTLPGPAMLLPELCYLCTGKMDKRNDFNMVKDL--AVHTRLTPQORQREVGR	337	
Db	1007	AEKRAKVESLQNKQILNPCLCAHPIPASLWRKAVCLPSILYRLHCLLTAEELRAQTAS	1066	
QY	338	-----LIDYIHKNDNVORELDWGLSFDPS--NLLSFS-----GRILQTE	374	
Db	1067	DAGVGVRSILPADFRYPNLDF-----GWKSIDSKSFISINSSSAENDNYCKSHSTIVPEN	1121	
QY	375	KTHGGKTFDYNPQPADWSKETRGAPLISVXP	-----LDNMLL	412
Db	1122	AAHOGANTSTLENHDMQSVNCR--TLTSESFGKLHVESADLTAINGLSYNQNLANGSY	1177	
QY	413	ITYTRNVEAANSLLIQLEPK---VTPAMGMQWRKAIMIEVDDE-----TEAYLRVLQ	460	
Db	1180	DLANRDFCQGNQL--NYKQEIFVQPTTSYSTONLYSYENQPSDECTLLSNKYLGDNA	1237	
QY	461	QKVTDATQIVVCLLSSNRKDYDAIKKYLCCTDPTPSCQCVARTLGKQOTVMAITKIAL	520	
Db	1238	NKSTSDGSPVMAVMPGT--TDTIQVLKGRMDSB-QSPSICYSSRTLGPNPGLLIQALTLIS-	1294	
QY	521	QNNCKMGELREVV---DIPKLVMIVGIDC-YHDMTAGRSIAGFVASINEGMTWFSR	575	
Db	1295	---NASDGNLERLEMGDSFLKHAITTVLCTYPDHAHEGRUSYM-----	1336	
QY	576	CIFQDRGQELVDGLKVLQALRAWNSCNEY-----MPSRIIV	613	
Db	1337	-----RSKKV-----SNCNLYELGKKGLPSEMVV	1361	

RESULT 12

DICE MOUSE

ID DICE MOUSE STANDARD; PRT; 1906 AA.

AC Q8R418;

CD 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Endoribonuclease Dicer (EC 3.1.26.-) (Double-strand-specific

DE ribonuclease mDicer-1)

GN DICER1 OR DICER OR MDCR.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI_TaxID=10090;

[1] _SEQUENCE FROM N.A.

RP MEDLINE=21886641; PubMed=11889553;

RX MEDLINE=21886641; PubMed=11889553;

RA Nicholson R.H., Nicholson A.W.;

RT "Molecular characterization of a mouse cDNA encoding Dicer, a

RL ribonuclease III ortholog involved in RNA interference.";

RL Mamm. Genome 13:67-73 (2002).

[2]

RP _SEQUENCE FROM N.A.

RP STRAIN=Czech II;

RC STRAIN=Czech II;

RA Svoboda P., Anger M., Stein P., Schultz R.M.;

RT "Mouse dicer homolog in oocyte and preimplantation embryos.";

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

[3]

RP _SEQUENCE FROM N.A.

RP Doi N., Zenko S., Ui-Tei K., Takahashi F., Ueda R., Miyata Y.,

RA Saigo K.;

RT "e12C family proteins and Dicer homologs are required for siRNA

RL mediated RNAi in mammalian cells.";

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: Involved in cleaving double-stranded RNA in the RNA

CC interference (RNAi) pathway. Itroduces 21 to 23 bp dsRNA_s

(siRNAs) which target the selective destruction of homologous RNAs.

CC - TISSUE SPECIFICITY: Expressed in a wide variety of tissues.

CC - SIMILARITY: BELONGS TO THE HELICASE FAMILY.

CC - SIMILARITY: Contains 2 RNase III domains.

CC - SIMILARITY: Contains 1 PAZ domain.

CC - SIMILARITY: Contains 1 DREM (double-stranded RNA-binding) domain.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF484524; AAL84638.1; -

DR EMBL; AF430845; AAM21495.1; ALT_INIT.

DR EMBL; AB081470; BAC15765.1; -

DR MGD; MGI:2177178; Dicer1.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR001159; DS_RBD.

DR InterPro; IPR005034; DUF283.

DR InterPro; IPR001650; Helicase_C.

DR InterPro; IPR003100; PAZ.

DR InterPro; IPR000999; RNase_3.

DR Pfam; PF002070; DEAD; 1.

DR Pfam; PF00035; DSRM; 1.

DR Pfam; PF03368; DUF283; 1.

DR Pfam; PF00271; helicase_C; 1.

DR Pfam; PF02170; PAZ; 1.

DR Pfam; PF00636; Ribonuclease_3; 2.

DR SMART; SM00487; DEXDC; 1.

DR SMART; SM00358; DSRM; 1.

DR SMART; SM00490; HELIC; 1.

DR SMART; SM00535; RIBOC; 2.

DR PROSITE; PS00137; DS_RBD; 1.

DR PROSITE; PS00821; PAZ; 1.

DR PROSITE; PS00517; RNase_3; 1; 1.

DR PROSITE; PS0142; RNase_3; 2; 2.

KW Helicase; ATP-binding; RNA-binding; Hydrolase; Nuclease; Endonuclease;

KW Repeat.

FT NP_BIND 34 41 ATP (POTENTIAL).

FT SITE 165 168 DECH BOX.

FT DOMAIN 881 1032 PAZ.

FT DOMAIN 1266 1393 RNase III 1.

FT DOMAIN 1650 1808 RNase III 2.

FT DOMAIN 1833 1898 DREM.

FT CONFLICT 97 97 A -> C (IN REF. 2).

FT CONFLICT 157 157 S -> P (IN REF. 3).

FT CONFLICT 279 279 H -> Y (IN REF. 2).

FT CONFLICT 600 600 A -> T (IN REF. 3).

FT CONFLICT 749 749 E -> D (IN REF. 2).

FT CONFLICT 827 827 T -> I (IN REF. 2).

FT CONFLICT 878 878 G -> S (IN REF. 1).

FT CONFLICT 955 955 Y -> C (IN REF. 3).

FT CONFLICT 983 983 T -> A (IN REF. 3).

FT CONFLICT 1080 1080 R -> G (IN REF. 3).

FT CONFLICT 1100 1100 T -> S (IN REF. 3).

FT CONFLICT 1326 1326 P -> H (IN REF. 1).

FT CONFLICT 1609 1609 A -> S (IN REF. 3).

FT CONFLICT 1850 1850 K -> E (IN REF. 3).

SQ SEQUENCE 1906 AA; 215766 MW; 230EA9BFC193091 CRC64;

Query Match 3.3%; Score 134.5; DB 1; Length 1906;

Best Local Similarity 19.3%; Pred. No. 0.23;

Matches 139; Conservative 90; Mismatches 277; Indels 213; Gaps 32;

QY 38 RPWALYQYHIDY---NPLMEARRLSALLFOHEDLIGKCHAPDGTILFLPKRLOQKVTVE 94

Db 753 KPDQPCYLVIGMVLATPLPDELNFRRLKLYPPEDTT-RCFG-----ILTAKPIQ-IPH 805

QY 95 VFSKTRNGEDVRITITITLWLPSTPTCLQ-----FNNIFRLLAKI-----MNLQOI 142

Db 806 FPVYTRSGE-----VTISIELKSGFTLSQOMLELITELHOYIFSHILREKPALEFKPT 860

QY 143 GR-----NYYNPNDPIDIPSHRLVWPGFTTSILOYENSIMLCSTDVSHKVRSTV 193

Db 861 GAESAYCVLPINLVNDSTGLDIFK-----FMEDIEKSEARIGIPSTKYSK-----ETP 909

QY 194 LDFMNFYHQTEEHKFOQVSKELIGLVLTYYNN-----KTYRVDDDDWDQNPSTKTEKA 249

Db 910 FVFKLEIDYQDA-----VILPRYRNFQOPHRFYADVYTDLTTPSKFESP 953

QY 250 DGSEVSFLEYRYKQYNOBITDLKQPVL-----VSPQKRRR 284

Db 954 EYE--TFAEYKYKYNLDLTNLNQPLLDVHTSSRLNLTTPRHLNQKCKALPSSAEK 1011

QY 285 GPGGTLPQFAMLIPELCVLTGLTKMRNDFNMKDL--AVHTRLTPEQOREVGR----- 337

Db 1012 AKWESLQNKQILVPELCAIHPASLWRKAVCLPSILYRLHCLLTAEELRAQTASDAGVG 1071

QY 338 ----LIDYIHKNDNVQRELDRWGLSFDG-----NLLSFGRILOTEKTHQ 378

Db 1072 VRSFVDFRYPNLDF-----GWKKSIDSFSFSTCNSSLAESDNYCKHSTTVVPHAAHQ 1126

QY 379 GG--KTFDYNQFADWSKETRGAPLISV-----KPLDNWLLIYTR 416

Db 1127 CATRPSLENHQD---SVNCKRLPAESPAKLOSEVSTDLTAINGLSYNKNLANGSYDLVN 1183

QY 417 RNYEAANSIQLNLFK---VTPAMGMQMKRAIMIEVDDR-----TEAYLRVLOQKVT 464

Db 1184 RDFCQGNQL--NYFKQELPVQFTTSYQIYNLYENQPKPSNECPLLSNTYLDGNANTST 1241

QY 465 ADTQIVVCLLSNR-----XKDYDAIKKYLCTDCPTSCQCVVARTLCKQQTWMAIATKI 518

Db 1242 SDGSPAVSTMPMAMNNAVKALDRMDSEQ-----SPSVGVSSRTLGNPGLILOALT 1293

QY 519 ALQWCKMKGGLWRV---DIPKLKLVMIIVGIDC-YHDMTAGRRSI----- 558

Db 1294 S---NASDGFNLERLEMLGDSFLKHAITYLFCYTPDAHEGRLSYMRSKVSCNLYRLG 1350

QY 559 -----AGFVASINEGWRWF--SRCIFQDRGQ-----ELVDGLKVCLOAALRAWNSCNE 605

Db 1351 KKKGLPSRMVVSIFDPPVNWLPVPGVYVNVQDKSNSEKWEKMDTKDCLLANGKLGEACEE 1409

RESULT 13

MFDP_RICPR

ID MFDP_RICPR STANDARD; PRT; 1120 AA.

AC 005955;

DT 15-DEC-1998 (Rel. 37, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Transcription-repair coupling factor (TRCF).

OS MFD OR RP598.

OS Rickettsia prowazekii.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

OC Rickettsiaceae; Rickettsiae; Rickettsia.

OX NCBI_TaxID=782;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Madrid E;

RX MEDLINE=99039499; PubMed=9823893;

RA Andersson S.G.E.; Zomorodipour A.; Andersson J.O.,

RA Sacheritz-Ponten T.; Alsmark U.C.M.; Podowski R.M.; Naeslund A.K.,

RA Eriksson A.-S.; Winkler H.H.; Kurland C.G.;

RT "The genome sequence of Rickettsia prowazekii and the origin of

RL mitochondria."

RL Nature 396:133-140(1998).

RN [2]

RP SEQUENCE OF 132-1120 FROM N.A.

RC STRAIN=Madrid E;

RX MEDLINE=97419517; PubMed=9274032;

RA Andersson J.O.; Andersson S.G.E.;

RT "Genomic rearrangements during evolution of the obligate

[illegible]

DR HSSP; P14687; 1AMU.
 DR InterPro; IPR000873; AMP-bind.
 DR InterPro; IPR001242; Condensatn.
 DR InterPro; IPR006163; Pp-bind.
 DR InterPro; IPR006162; Ppantne attach.
 DR InterPro; IPR000511; SAM bind.
 DR Pfam; PF00501; AMP-binding; 1.
 DR Pfam; PF00668; Condensation; 2.
 DR Pfam; PF00550; pp-binding; 3.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
 DR PROSITE; PS00455; AMP BINDING; 1.
 DR PROSITE; PS00075; ACP DOMAIN; 3.
 KW Ligase; Multifunctional enzyme; Phosphopantetheine; Repeat.
 FT DOMAIN 3 547
 FT REPEAT 114 146
 FT REPEAT 310 321
 FT REPEAT 378 390
 FT REPEAT 454 462
 FT REPEAT 477 491
 FT DOMAIN 1466 1919
 FT REPEAT 1495 1527
 FT REPEAT 1682 1693
 FT REPEAT 1750 1762
 FT REPEAT 1826 1834
 FT REPEAT 1849 1863
 FT DOMAIN 20 88
 FT DOMAIN 1409 1475
 FT DOMAIN 1944 2014
 FT BINDING 52 52
 SQ SEQUENCE 2035 AA; 228826 MW; 1C801377A4375BDC CRC64;
 Query Match 2.8%; Score 114; DB 1; Length 2035;
 Best Local Similarity 18.3%; Pred. No. 7.9;
 Matches 150; Conservative 124; Mismatches 327; Indels 220; Gaps 38;
 QY 25 IYRLSTNHFLRSPQWALYQYHIDYINPLMEARRLSALLFQH-----EDLIKCHAF 77
 DB 1129 ILRGARNELSLNDAQS-----PESLAFNPASAPYIOELATICOQ- 1170
 QY 78 DGTILFLPRLOQKTEVFSKTEKNGEDVITITLINELPPTCTCLOFNIIFRLLKIM 137
 DB 1171 -----LAQRLQRPVLELVEGTGTGRAESLLAQLN-----AGQIEYV 1207
 QY 138 NLQIGRNNYNNPDIDIFSHRLVIWPGF-----TTSILQYENSIMLCITDVSHKVL 189
 DB 1208 GLEQ-----SCEMLLSARQLASPGARLSPWADTLAAHSHGDIILWNLALHRLP 1260
 QY 190 SE-----TVLDFMNFYHYOTEHKFQVSKELIGLVLTLYKNNKTYRVD 235
 DB 1261 EDPGLLATIQLQAVPGALLYVM-----EFRQLTSPALLSTLLTNGQPEALLHNS 1310
 QY 236 IDWDQNPKSTFKKADGSEVSYFLEYRKYNOYNOETDLKQPVLSQPKR--RRGP----- 286
 DB 1311 ADW-----AALFSA-----AFNCQHSDEVAGL-QRFLVQCPRQVRERDPRQLQAL 1356
 QY 287 GGTLPGPAMLIPE-LCYLGT--LTDKMRDNFNMKDLAVHTRLTTPQR----- 331
 DB 1357 AGRLPG--WVVPQRIYFLDALPLTANGKIDYQALK-----RHTPKAENQAEADLPQGL 1409
 QY 332 QREVGELDIYHKNDNVQRELWDGLSFDNSLLSFGSRILOTEKHGGKTFDYNQPAD 391
 DB 1410 EKQVAALWQQLSTGNVTRTDFPQGGDSLLAT--RL--TGQLHOAG---YEQLSD 1460
 QY 392 WSKETGAPLISV-----KPLDNWLLIYTRNEYA-ANSLIQMLFKVTYAMGQMKRA-- 443
 DB 1461 LFNHPRADFAATLRKIDVPVFPFHSPEERYQFPALTDVQVQAYLVGRPGFTLGGVS 1520
 QY 444 ---IMLEVDRTAYRLVQKVTADTQIVVCLLSNRKDKYDAIKYLCTDCTPTSCQV 500
 DB 1521 HFFVEFEIADLDLTRLETVMNRILAHMDLRVAVLDGQQQVLE-----QTFPFWI 1570
 QY 501 VARTLGKQQTVMIAIKALQNMCKMGELWRY-DIPLKLVIMVIGDICVHDMTAGRSIA 559

DB 1571 PTHLTPEEARLVREKLAHQV--LNPVWPVFDLQV----- 1605
 QY 560 GFVASINEGTRWFSRCIFQDRGOELVDGL--KYCLOAALRAMNSCMEYPSRIIVYRDG 617
 DB 1606 GIV-----DGMARLWLCI-----DNLLDGLSMQILLAELEGYRYPQQLLPPLVTFRDY 1657
 QY 618 VGDQLKTLVNYEVPQFQLDCLKISIGRYNPRLTIVVVKRYNTRFFAOSGGRLQNP----- 673
 DB 1658 LQPSLSQSPNEDSLAWQAQLDDIPPA--PALPRLCLPQEVETPRFARLNGALDSTRWHR 1715
 QY 674 LPGTVIVDVETRPEDWYFFIVSQAVRSGVSP--THYNYIYDNGSLKPDHQLRLTYKLCH 731
 DB 1716 LKRAADAHLT-PSAVLLSVWSTVLSAWSAQPEFTLNLTLFDRPLHPQINQILGDTSL 1774
 QY 732 IYYNW-PGVIRVPAPCOYAHKLAFLVGSIHREPNLSLSNR 771
 DB 1775 MLLSWHGE-----SWLHS-----AQSLQORLSQNLNR 1803
 RESULT 15
 PHAC_PSEOL STANDARD; PRT; 560 AA.
 ID PHAC_PSEOL
 AC P26496;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Poly(3-hydroxyalkanoate) polymerase 2 (EC 2.3.1.-) (PHA polymerase 2)
 DE (PHA synthase 2) (Polyhydroxyalkanoic acid synthase 2).
 GN PHAC.
 OS Pseudomonas oleovorans.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=301;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GP01;
 RX MEDLINE=91115830; PubMed=1989978;
 RA Huisman G.W., Woonink E., Meima R., Kazemier B., Terpstra P.,
 RA Witholt B.;
 RT "Metabolism of poly(3-hydroxyalkanoates) (PHAs) by Pseudomonas
 RT oleovorans. Identification and sequences of genes and function of the
 RT encoded proteins in the synthesis and degradation of PHA.";
 RL J. Biol. Chem. 266:2191-2198(1991).
 CC -!- FUNCTION: P.OLEOVORANS ACCUMULATES POLY(3-HYDROXYALKANOATES) AFTER
 CC GROWTH ON MEDIUM CHAIN LENGTH HYDROCARBONS. LARGE AMOUNTS OF THIS
 CC POLYESTER ARE SYNTHESIZED WHEN CELLS ARE GROWN UNDER NITROGEN-
 CC LIMITING CONDITIONS. WHEN NITROGEN IS RESUPPLIED IN THE MEDIUM,
 CC THE ACCUMULATED PHA IS DEGRADED.
 CC -!- SIMILARITY: BELONGS TO THE PHA/PHB SYNTHASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; M58445; AA25934.1; -.
 DR PIR; C38604; C38604.
 DR InterPro; IPR000073; A/b hydrolase.
 DR Pfam; PF00561; abhydrolase; 1.
 KW PHA biosynthesis; Transferrase; Acyltransferase.
 FT ACT_SITE 296 296 POTENTIAL.
 SQ SEQUENCE 560 AA; 62631 MW; E2CD844FC1616B83 CRC64;
 Query Match 2.8%; Score 113.5; DB 1; Length 560;
 Best Local Similarity 19.6%; Pred. No. 1.3;
 Matches 119; Conservative 80; Mismatches 231; Indels 177; Gaps 27;
 QY 164 PGFTTSLILOYENSIMLCITDVSHKVLSRSETVLDPMFNFYHQTEHKFQE-----QVS 214
 DB 10 PTLPATSMNVQNAI-----GLGRDLISLNRVSRQSLRHLPLHTAHLALGGQLG 61

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 6, 2003, 17:51:05 ; Search time 4855 Seconds
(without alignments)
11654.134 Million cell updates/sec

Title: US-10-043-774B-1
Perfect score: 2328
Sequence: 1 atgatctttggtgaacac.....accgcctttactacctctaa 2328

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hic:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estcom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vit:*

21: em_gss_fun:*

22: em_gss_man:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pig:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	484	19.5	527	12	BM751055 K-EST0026
2	452	19.4	874	10	BF666315
3	452	19.4	1495	10	BF664941
4	413	17.7	413	9	AL041473

5	405	17.4	642	13	BU854142
6	361	15.5	587	9	AA430311
7	321	13.8	462	9	AA969938
8	318	13.7	779	12	BI560710
9	308	13.2	818	12	EG773137
10	285	12.2	622	9	AL705072
11	275	11.8	559	12	BM728018
12	257	11.0	420	13	BQ374910
13	249	10.7	718	13	BQ435882
14	245	10.5	512	12	EM680592
15	241	10.4	640	9	AL705301
16	241	10.4	1092	12	BI463482
17	232	10.0	378	10	BF764928
18	226	9.7	226	10	BF092403
19	223	9.6	731	12	BI458664
20	218	9.4	715	10	EG721550
21	215	9.2	581	10	EG718168
22	208	8.9	705	10	BF246708
23	195	8.4	226	10	BF092519
24	192	8.2	605	9	AL704986
25	179	7.7	539	12	BI463615
26	178	7.6	666	12	BI463897
27	164	7.0	807	13	BUI60046
28	125	5.4	522	9	AA456921
29	123	5.3	882	12	BI459464
30	108	4.6	470	13	EX281582
31	99	4.3	474	9	AL705611
32	91	3.9	830	10	BF699368
33	83	3.6	533	10	BF080923
34	83	3.6	577	10	BF080918
35	75	3.2	806	10	EG724366
36	71	3.0	744	12	BI459142
37	68	2.9	101	10	BF092393
38	60	2.6	113	13	EQ374884
39	57	2.4	935	10	BF247005
40	54	2.3	228	13	BQ375179
41	41	1.8	485	10	EG718140
42	35	1.5	477	12	BM539361
43	35	1.5	565	10	EG101107
44	34	1.5	872	12	BI560401
45	21	0.9	348	13	BY018987

ALIGNMENTS

RESULT 1
BM751055
LOCUS
K-EST0026956 S9SNU601 Homo sapiens cDNA clone S9SNU601-21-E08 5', linear EST 04-MAR-2002
DEFINITION
mRNA sequence.
ACCESSION
BM751055
VERSION
BM751055.1 GI:19080673
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. (bases 1 to 527)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.
21C Frontier Korean EST Project 2001
JOURNAL
Unpublished
COMMENT
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 21 row: E column: 08
High quality sequence stop: 527.


```

Db 310 CGATTGTTACCATGATGACAGCTGGCGGAGTCAATCGCAGGATTGTTGCCAGCAT 369
QY 1695 CAATGAGGATGACCCGCTGTTCTCAGCTGCATATTTACGATAGAGGACAGAGCT 1754
Db 370 CAATGAGGATGACCCGCTGTTCTCAGCTGCATATTTACGATAGAGGACAGAGCT 429
QY 1755 GGTAGATGGCTCAAAAGTCTGCTGCAAGCGG 1786
Db 430 GGTAGATGGCTCAAAAGTCTGCTGCAAGCGG 461

```

```

RESULT 3
BF664941 1495 bp mRNA linear EST 21-DEC-2000
LOCUS 602119191F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4276247 5',
DEFINITION mRNA sequence.

```

```

ACCESSION BF664941
VERSION BF664941.1 GI:11938836
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 1495)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC

```

```

CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1096 row: f column: 24
High quality sequence stop: 511.
Location/Qualifiers
1..1495
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4276247"
/tissue_type="primitive neuroectoderm"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 56"
/notes="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggcgctcgcc); Site_2: SfiI (ggccattatggcc);
adaptor sequence: 5'-CACGCCATTATGCGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGCGCGACATG-DT(30)BN-3'
(where B = A, C, G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
```

```

FEATURES
source

```

```

BASE COUNT 476 a 280 c 536 g 203 t
ORIGIN
Query Match 19.4%; Score 452; DB 10; Length 1495;
Best Local Similarity 100.0%; Pred. No. 5.9e-235;
Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1335 GATTGAAGTGATGACAGAACTCAAGCTACTTAAGAGTCTTACAGCAAAAGTCAACG 1394
Db 10 GATTGAAGTGATGACAGAACTCAAGCTACTTAAGAGTCTTACAGCAAAAGTCAACG 69
QY 1395 AGACACCCAGATAGTTGTCTGTCTGTGTCAAGTAATCGAAGGACAAATACGATCTAT 1454
Db 70 AGACACCCAGATAGTTGTCTGTCTGTGTCAAGTAATCGAAGGACAAATACGATCTAT 129

```

```

QY 1455 TAAAAATAACTGTGTACAGATTGGCCCTACCCCAAGTCAGTGTGTGTGGCCGACCTT 1514
Db 130 TAAAAATAACTGTGTGTACAGATTGGCCCTACCCCAAGTCAGTGTGTGTGGCCGACCTT 189
QY 1515 AGGCAAAACAGCAAACTGTTCATGGCCATTGCTACAAAGATTGGCCCTACAGATGAACCTGCAA 1574
Db 190 AGGCAAAACAGCAAACTGTTCATGGCCATTGCTACAAAGATTGGCCCTACAGATGAACCTGCAA 249
QY 1575 GATGGAGGAGAGCTCTCTGGAGGTGGACATCCCTCCCTGAAGCTCGTCGATGATCGTTGGCAT 1634
Db 250 CATGGAGGAGAGCTCTCTGGAGGTGGACATCCCTCCCTGAAGCTCGTCGATGATCGTTGGCAT 309
QY 1635 CGATTGTTACCATGACATCAGCTGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCAT 1694
Db 310 CGATTGTTACCATGACATCAGCTGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCAT 369
QY 1695 CAATGAAGGATGACCCGCTGTTCTCAGCTGTCATATTTACGATAGAGGACAGGAGCT 1754
Db 370 CAATGAAGGATGACCCGCTGTTCTCAGCTGTCATATTTACGATAGAGGACAGGAGCT 429
QY 1755 GGTAGATGGCTCAAAAGTCTGCTGCAAGCGG 1786
Db 430 GGTAGATGGCTCAAAAGTCTGCTGCAAGCGG 461

```

```

RESULT 4
AL041473/c
LOCUS
DEFINITION

```

```

AL041473 413 bp mRNA linear EST 29-FEB-2000
DEFINITION DKFZp434E2417.s1.434 (synonym: htes3) Homo sapiens cDNA clone
DKFZp434E2417.3, mRNA sequence.

```

```

ACCESSION AL041473.1 GI:5420824
VERSION
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 413)
AUTHORS Bloecker, H., Boecker, M., Brandt, P., Mewes, H.W., Gassenhuber, J. and
Wiemann, S.
TITLE EST (Bloecker, et al.)
JOURNAL Unpublished
COMMENT Contact: Bloecker H
MIPS

```

```

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 3' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.

```

```

ri sequence also available.
This clone (DKFZp434E2417) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

```

```

FEATURES
source

```

```

1..413
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp434E2417"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="434 (synonym: htes3)"
/notes="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT 100 a 118 c 83 g 112 t
ORIGIN

```

```

Query Match 17.7%; Score 413; DB 9; Length 413;
Best Local Similarity 100.0%; Pred. No. 8.5e-214;
Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1637 ATTGTTACCATGACATGACAGCTGGGCGAGAGTCAATCGCAGGATTTGTTCCAGCATCA 1696
 Db |||||
 QY 413 ATTGTTACCATGACATGACAGCTGGGCGAGAGTCAATCGCAGGATTTGTTCCAGCATCA 354
 Db |||||
 QY 1697 ATGAAGGATGACCGCTGGTTCTCAGCTGCATATTTCCAGTAGAGGACAGAGCTGG 1756
 Db |||||
 QY 353 ATGAAGGATGACCGCTGGTTCTCAGCTGCATATTTCCAGTAGAGGACAGAGCTGG 294
 Db |||||
 QY 1757 TAGATGGGCTCAAAAGTCTGCTCAAGCGGCTCTGAGGGCTTGAATAGCTGCAATGAGT 1816
 Db |||||
 QY 293 TAGATGGGCTCAAAAGTCTGCTCAAGCGGCTCTGAGGGCTTGAATAGCTGCAATGAGT 234
 Db |||||
 QY 1817 ACATGCCCGCGGATCATCTGTACCGCGATGCGGTAGAGAGCGCGCAGCTGAAAACAC 1876
 Db |||||
 QY 233 ACATGCCCGCGGATCATCTGTACCGCGATGCGGTAGAGAGCGCGCAGCTGAAAACAC 174
 Db |||||
 QY 1877 TGGTGAACACTAGAGTGGCCAGCTTTTGGATTGCTAAATCATTTGGTAGAGTTACA 1936
 Db |||||
 QY 173 TGGTGAACACTAGAGTGGCCAGCTTTTGGATTGCTAAATCATTTGGTAGAGTTACA 114
 Db |||||
 QY 1937 ACCCTAGACTAACGGTAATTGCTGTGAAGAAAGAGTGAACACAGATTTTTCCTCAGT 1996
 Db |||||
 QY 113 ACCCTAGACTAACGGTAATTGCTGTGAAGAAAGAGTGAACACAGATTTTTCCTCAGT 54
 Db |||||
 QY 1997 CTGAGGAGAGCTTCAGATCCACTCTCTGAAACAGTATTGATGATGAGGTT 2049
 Db |||||
 QY 53 CTGAGGAGAGCTTCAGATCCACTCTCTGAAACAGTATTGATGATGAGGTT 1

RESULT 5
 BU854142
 LOCUS
 DEFINITION BU854142 642 bp mRNA linear EST 16-OCT-2002
 AGENCOURT_10402932 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6621314
 5', mRNA sequence.
 ACCESSION BU854142
 VERSION BU854142.1 GI:24039108
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 642)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: CLONTECH
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Cloned by: Agencourt Bioscience Corporation
 Cloning distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LCM2872 row: n column: 02
 High quality sequence stop: 360.

FEATURES
 source
 1. .642
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6621314"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NIH_MGC_82"
 /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site: 1:
 SfiI (ggcgctcgcc); Site 2: SfiI (ggcattatggcc); 5' and
 3' adaptors were used in cloning as follows: 5' adaptor
 sequence: 5'-CACGGCCATATGGCC-3', and 3' adaptor sequence:
 5'-ATTCTAGAGCGAGGCGGCGACATG-dt(30)BN-3' (where B = A,
 C, or G and N = A, C, G, or T). Average insert size
 1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained
 inserts by PCR. This library was enriched for full-length

clones and was constructed by Clontech Laboratories (Palo
 Alto, CA)."
 BASE COUNT 194 a 138 c 177 g 133 t
 ORIGIN
 Query Match 17.4%; Score 405; DB 13; Length 642;
 Best Local Similarity 100.0%; Pred. No. 2.3e-209;
 Matches 405; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1450 GCTATTAAAAATACCTGTCTACAGATTGCCCTACCCCAAGTCAGTGTGTGGCCCGA 1509
 Db GCTATTAAAAATACCTGTCTACAGATTGCCCTACCCCAAGTCAGTGTGTGGCCCGA 62
 QY 1510 ACCTTAGGCAAAACAGCAAACTGTCTATGGCCATTGTCTACAAAGATTGCCCTACAGATGAAC 1569
 Db ACCTTAGGCAAAACAGCAAACTGTCTATGGCCATTGTCTACAAAGATTGCCCTACAGATGAAC 122
 QY 1570 TGAAGATGGAGGAGAGCTCTGAGGGTGGACATCCCTGAAAGCTCGTGTGATGATGTT 1629
 Db TGAAGATGGAGGAGAGCTCTGAGGGTGGACATCCCTGAAAGCTCGTGTGATGATGTT 182
 QY 1630 GGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTGTTGCC 1689
 Db GGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTGTTGCC 242
 QY 1690 AGCATCAATGAAGGATGACCCGCTGTTCTACGCTGCATATTTTCAGGATAGAGACAG 1749
 Db AGCATCAATGAAGGATGACCCGCTGTTCTACGCTGCATATTTTCAGGATAGAGACAG 302
 QY 1750 GAGCTGGTAGAGGCTCAAAAGTCTGCTGCAAGCGGCTCTGAGGGCTTGGATAGTGC 1809
 Db GAGCTGGTAGAGGCTCAAAAGTCTGCTGCAAGCGGCTCTGAGGGCTTGGATAGTGC 362
 QY 1810 AATGAGTACATGCGCCAGCGGATCATCGTGTACCGCGATGGCGTA 1854
 Db AATGAGTACATGCGCCAGCGGATCATCGTGTACCGCGATGGCGTA 407

RESULT 6
 AA430311 587 bp mRNA linear EST 16-OCT-1997
 LOCUS
 DEFINITION 2w68h01.r1 Soares testis_NHT Homo sapiens cDNA clone IMAGE:781393
 5', similar to WP:00105.2 CE06748 ;, mRNA sequence.
 ACCESSION AA430311
 VERSION AA430311.1 GI:2113521
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 587)
 AUTHORS Hillier, L., Allen, M., Bowles, L., Dubucque, T., Geisel, G., Jost, S.,
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,
 Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie,
 T., Waterston, R., and Wilson, R.
 WashU-Merck EST Project 1997
 TITLE JOURNAL
 COMMENT Unpublished
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28m3 rev2 ET from Amersham
 High quality sequence stop: 498.
 Location/Qualifiers
 1. .587
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:781393"

```

/sex="male"
/lab_host="DH10B"
/clone_lib="Soares testis NHT"
/notes="Vector: pMT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, inc., and primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAAGTGGAGCGCGCCCAATTTTGTGATTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pMT3D vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 191 a 122 c 136 g 138 t
ORIGIN

Query Match 15.5%; Score 361; DB 9; Length 587;
Best Local Similarity 100.0%; Pred. No. 2.7e-185;
Matches 361; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1852 GTAGGACGCGCCAGCTGAAACACTGCTGAACCTACGAGTGCACAGTTTGGATTGT 1911
Db 38 GTAGGACGCGCCAGCTGAAACACTGCTGAACCTACGAGTGCACAGTTTGGATTGT 97
QY 1912 CTAATATCCATTGTTAGAGTTTACACCCCTAGACTTAACCGGTAAATGTTGGTGAAGAAAGA 1971
Db 98 CTAATATCCATTGTTAGAGTTTACACCCCTAGACTTAACCGGTAAATGTTGGTGAAGAAAGA 157
QY 1972 GTGACACCAAGATTTTGTCTAGTCTGAGGAGAAAGACTTCAGAAATCCATTCCTGGGAACA 2031
Db 158 GTGACACCAAGATTTTGTCTAGTCTGAGGAGAAAGACTTCAGAAATCCATTCCTGGGAACA 217
QY 2032 GTTATTGATGTAGAGTTTACACACCAAGATGTTATGATGATGATGATGATGATGATGATGAT 2091
Db 218 GTTATTGATGTAGAGTTTACACACCAAGATGTTATGATGATGATGATGATGATGATGATGAT 277
QY 2092 GTGAGAAAGTGTAGTGTGTTCTCCACACATTACAAATGTCATCTATGATGATGATGATGATGAT 2151
Db 278 GTGAGAAAGTGTAGTGTGTTCTCCACACATTACAAATGTCATCTATGATGATGATGATGATGAT 337
QY 2152 AAGCCAGACACATACAGCGCTTGACCTACAGTGTGCCACATCTATTACAACTGSCCA 2211
Db 338 AAGCCAGACACATACAGCGCTTGACCTACAGTGTGCCACATCTATTACAACTGSCCA 397
QY 2212 G 2212
Db 398 G 398

RESULT 7
AA969938 462 bp mRNA linear EST 07-JUL-1998
LOCUS OP60a11.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:1581212.3 similar to WP:00165.2 CB06748 ;, mRNA sequence.
ACCESSION AA969938
VERSION AA969938.1 GI:3145451
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 462)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 521 Std Error: 0.00
Seq primer: -40m13 fwd. RT from Amersham.
Location/Qualifiers

BASE COUNT 127 a 97 c 124 g 114 t
ORIGIN

Query Match 13.8%; Score 321; DB 9; Length 462;
Best Local Similarity 99.8%; Pred. No. 1.9e-163;
Matches 441; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1714 TGGTTCTCAGCTGCATATTTTCCAGGATAGAGACAGGAGCTGGTAGATGGCTCAAAGTC 1773
Db 1 TGGTTCTCAGCTGCATATTTTCCAGGATAGAGACAGGAGCTGGTAGATGGCTCAAAGTC 60
QY 1774 TGCCTCAAGCGGCTCTGAGGGCTTGGAAATAGCTCAATGAGTACATGCCAGCCGATC 1833
Db 61 TGCCTCAAGCGGCTCTGAGGGCTTGGAAATAGCTCAATGAGTACATGCCAGCCGATC 120
QY 1834 ATCGTGTACCGGATGCGCTAGGAGCGGCGCTCAAAACACTGTGTGAATACGAGTG 1893
Db 121 ATCGTGTACCGGATGCGCTAGGAGCGGCGCTCAAAACACTGTGTGAATACGAGTG 179
QY 1894 CACAGTTTTTGGATTGTTCTAAATCCATGTTAGAGTTTCAACCCCTAGACTTAACGGTA 1953
Db 180 CACAGTTTTTGGATTGTTCTAAATCCATGTTAGAGTTTCAACCCCTAGACTTAACGGTA 239
QY 1954 ATTGTGCTGAAGAAAGAGTGAACACAGATTTTTCCTCAGTCTGAGGAGAACTTCAG 2013
Db 240 ATTGTGCTGAAGAAAGAGTGAACACAGATTTTTCCTCAGTCTGAGGAGAACTTCAG 259
QY 2014 AATCCACTTCTGGAACAGTTATTGATGTAGAGTTTACCAGACCAAGATGGTATGACTTT 2073
Db 300 AATCCACTTCTGGAACAGTTATTGATGTAGAGTTTACCAGACCAAGATGGTATGACTTT 359
QY 2074 TTTATGCTGAGCCAGCTGTGAGAGTGTGTTCTTCCACACATTACATGATGATC 2133
Db 360 TTTATGCTGAGCCAGCTGTGAGAGTGTGTTCTTCCACACATTACATGATGATC 419
QY 2134 TATGACAAACAGCGGCTGGAAGC 2155
Db 420 TATGACAAACAGCGGCTGGAAGC 441

RESULT 8
BI560710 779 bp mRNA linear EST 05-SEP-2001
LOCUS 603254744.F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5296954.5',
DEFINITION mRNA sequence.
ACCESSION BI560710
VERSION BI560710.1 GI:15448024
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 779)
NIH-MGC http://mgc.ncbi.nih.gov/.
AUTHORS

```



```

VERSION AL705072.1 GI:19688427
SOURCE EST.
ORGANISM Homo sapiens (human)
REFERENCE Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 622)
Ottawaelder, B., Obermaier, B., Mewes, W., Mewes, H.W., Weil, B. and
Wiemann, S.
TITLE EST (Ottawaelder, B., Obermaier, B., Mewes, H.W., Weil, B. and Wiemann
S.)
JOURNAL Unpublished
COMMENT Contact: Ottawaelder B
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by MedGenomix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project. No sl sequence
available.
This clone (DKFZp686E0634) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
Location/Qualifiers
1..622
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686E0634"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: blcc3)"
/note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"
BASE COUNT 187 a 134 c 168 g 133 t
ORIGIN
Query Match 12.2%; Score 285; DB 9; Length 622;
Best Local Similarity 100.0%; Pred. No. 16-143;
Matches 285; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 32 TAGACCATGTTAAAGATCAAAACAGGTTCTTCAGGCATTATAGTAAGTTAAGCACTA 91
Db 338 TAGACCATGTTAAAGATCAAAACAGGTTCTTCAGGCATTATAGTAAGTTAAGCACTA 397
QY 92 ACCATTTCGGCTGACATCCCGTCCCGAGTGGCCCTTATATCATGATTCATTCAGTATA 151
Db 398 ACCATTTCGGCTGACATCCCGTCCCGAGTGGCCCTTATATCATGATTCATTCAGTATA 457
QY 152 ACCCATCTGATGGAAGCAGAGACTCCGTTTCAGCTCTCTTTTCAACGAGAGTCTAA 211
Db 458 ACCCATCTGATGGAAGCAGAGACTCCGTTTCAGCTCTCTTTTCAACGAGAGTCTAA 517
QY 212 TTGGAAGTGTCTATGCTTTTGTAGGAACGATATTTTACCTAAAGACTACACAAA 271
Db 518 TTGGAAGTGTCTATGCTTTTGTAGGAACGATATTTTACCTAAAGACTACACAAA 577
QY 272 AGTTACTGAAGTTTATAGTAAGCCGGAATGGAGAGATGGA 316
Db 578 AGTTACTGAAGTTTATAGTAAGCCGGAATGGAGAGATGGA 622
RESULT 11
BM728018
LOCUS 559 bp mRNA linear EST 01-MAR-2002
DEFINITION UI-E-EJO-ait-h-19-0-UI.r1 UI-E-EJO Homo sapiens cDNA clone
BM728018
ACCESSION UI-E-EJO-ait-h-19-0-UI 5', mRNA sequence.
VERSION BM728018.1 GI:19049351
KEYWORDS EST.
SOURCE Homo sapiens (human)

```

```

ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 559)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalisation and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
889548
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.
FEATURES
Location/Qualifiers
1..559
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-EJO-ait-h-19-0-UI"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-E-EJO"
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EJO is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAAGA
; lens, CGATTAGCGA; eye anterior segment, AATGCCCAT;
optic nerve, CCATTAAATG; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
BASE COUNT 170 a 114 c 144 g 130 t
ORIGIN
Query Match 11.8%; Score 275; DB 12; Length 559;
Best Local Similarity 100.0%; Pred. No. 2.9e-138;
Matches 275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1938 CCTAGACTAACGGTAATGTGGTGAAGAAAGAGTGAACACAGATTTTGTCTCAGTC 1997
Db 212 CCTAGACTAACGGTAATGTGGTGAAGAAAGAGTGAACACAGATTTTGTCTCAGTC 271
QY 1998 TGGAGGAAGACTTCAGATCCACTTCCTGGACACTTATGATGATGAGTTACAGACC 2057
Db 272 TGGAGGAAGACTTCAGATCCACTTCCTGGACACTTATGATGATGAGTTACAGACC 331
QY 2058 AGAATGGTATGACTTTTATCGTAGCCAGCTCTGTGAGAGTGTAGTTCTCCAC 2117
Db 332 AGAATGGTATGACTTTTATCGTAGCCAGCTCTGTGAGAGTGTAGTTCTCCAC 391
QY 2118 ACATTACAATGTCATCTATGACACAGCGCTGAGCCAGCCACACATACAGCGCTTGAC 2177

```

```

Db      392 ACATTCAATGTTCATCTATGACAAACGCGCCCTGAAGCCACACATACAGCGCTTGAC 451
      |||||||
QY      2178 CTACAAAGCTGTGCCACATCTATTACAACTGGCCAG 2212
      |||||||
Db      452 CTACAAAGCTGTGCCACATCTATTACAACTGGCCAG 486
      |||||||

RESULT 12
BQ374910
LOCUS      420 bp mRNA linear EST 21-MAY-2002
DEFINITION MR4-TN0109-060900-101-h03 TN0109 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BQ374910
VERSION    BQ374910.1 GI:21050424
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 420)
          Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
          Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
          Goldman,G.H., Carvalho,A.P., Matsukuma,A., Baia,G.S., Simpson,D.H.,
          Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
          M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J.
          Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
          Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
          20202663
          10737800
          Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR4&t2=MR4-TN0109-060900-101-h03&t3=2000-09-08&t4=1)
          Seq primer: puc 18 forward
          High quality sequence stop: 420.
          Location/Qualifiers
            1..420
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /dev_stage="Adult"
              /clone_lib="TN0109"
              /notes="Organ: testis normal; Vector: puc18; Site 1: SmaI;
              Site 2: SmaI; A mini-library was made by cloning products
              derived from OPRESTES PCR (U.S. Letters Patent application
              No.196,716 - Ludwig Institute for Cancer Research)
              profiles into the pUC 18 vector. Reverse transcription of
              tissue mRNA and cDNA amplification were performed under
              low stringency conditions."
          BASE COUNT      122 a 94 c 100 g 104 t
          ORIGIN
          Query Match      11.0%; Score 257; DB 13; Length 420;
          Best Local Similarity 99.7%; Pred. No. 1.9e-128;
          Matches 307; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1905 GGATTGCTCTAAATCCATTCGTAGAGGTACACCCCTAGACTACACCGTAATTCGTGTGAA 1964
      |||||||
Db      97 GGATTGCTCTAAATCCATTCGTAGAGGTACACCCCTAGACTACACCGTAATTCGTGTGAA 156
      |||||||

QY      1965 GAAAGAGTGAACACACAGATTTTTCCTAGTCTGGAGGAGACTTCAGATCCATTC 2024
      |||||||
Db      157 GAAAGAGTGAACACACAGATTTTTCCTAGTCTGGAGGAGACTTCAGATCCATTC 216
      |||||||

```

```

QY      2025 TCGAACAGTATTGATGTAGAGTTTACAGACCAAGATGGTATGACTTTTATCTGAG 2084
      |||||||
Db      217 TCGAACAGTATTGATGTAGAGTTTACAGACCAAGATGGTATGACTTTTATCTGAG 276
      |||||||
QY      2085 CCAGGCTGTGAGAAGTGGTGTGTTCTCCACACATTACATGTCATCTATGACACAG 2144
      |||||||
Db      277 CCAGGCTGTGAGAAGTGGTGTGTTCTCCACACATTACATGTCATCTATGACACAG 336
      |||||||
QY      2145 CGGCTCTGAAGCCAGACCAATACATACAGCGCTTGACCTCAAGCTGTGCCACATCTATTACAA 2204
      |||||||
Db      337 CGGCTCTGAAGCCAGACCAATACATACAGCGCTTGACCTCAAGCTGTGCCACATCTATTACAA 396
      |||||||
QY      2205 CTGGCCAG 2212
      |||||||
Db      397 CTGGCCAG 404
      |||||||

RESULT 13.
BQ435882
LOCUS      718 bp mRNA linear EST 24-MAY-2002
DEFINITION AGENCOURT 7897283 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6103498
          5', mRNA sequence.
ACCESSION  BQ435882
VERSION    BQ435882.1 GI:21174947
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 718)
          NIH-MGC http://mgc.ncl.nih.gov/.
          National Institutes of Health, Mammalian Gene Collection (MGC)
          Unpublished
          Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          Tissue Procurement: CLONTECH
          cDNA Library Preparation: CLONTECH Laboratories, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLCM2339 row: n column: 11
          High quality sequence stop: 518.
          Location/Qualifiers
            1..718
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:6103498"
              /lab_host="DH10B (T1 phage-resistant)"
              /clone_lib="NIH MGC 82"
              /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site 1:
              SfiI (ggccattagccc); Site 2: SfiI (ggccattagccc); 5' and
              3' adaptors were used in cloning as follows: 5' adaptor
              sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence:
              5'-ATTCTAGAGCGCGCGCGACATG-dt(30)BN-3' (where B = A,
              C, or G and N = A, C, G, or T). Average insert size
              1.35 kb (range 0.9-4.0 kb). 14/15 colonies contained
              inserts by PCR. This library was enriched for full-length
              clones and was constructed by Clontech Laboratories (Palo
              Alto, CA)."
          BASE COUNT      192 a 137 c 140 g 249 t
          ORIGIN
          Query Match      10.7%; Score 249; DB 13; Length 718;
          Best Local Similarity 100.0%; Pred. No. 5.3e-124;
          Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2080 GTGACCCAGGCTGTGAGAAGTGGTATGTTTCTCCACACATTACATGTCATCTATGAC 2139
      |||||||
Db      5 GTGACCCAGGCTGTGAGAAGTGGTATGTTTCTCCACACATTACATGTCATCTATGAC 64
      |||||||

```

first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA; lens, CGAATAGCGA; eye anterior segment, AATGCCGGCAT; optic nerve, CATTAAGTGT; retina, CGCGG; Retina Poveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).

TAG_TISSUE=Foveal and Macular Retina

Query Match 10.5%; Score 245; DB 12; Length 512;

Best local similarity 100.0%, Fred. NO. 7.5e-122;
Matches 245; Conservative 0; Mismatches 0; Indels

1938 CCCTAGACTAACGGTAAATTGTGGTGAAGAAAAGAGTGCAACACACAGATTTTT

344 CCCTAGACTAACGGTAA^{TT}GTGGTGAAGAAAAGAGTGAA^{CAC}CACCA^{GA}TTT^{TTT}

1998 TGGAGGAAGACTTCAGATCCACTTCCTGGAACAGTTATTGATGTAGAGGT

284 TGGAGGAAGACTTCAGAAATCCACTTCTCGAACAGTTATTGATGTAGAGGT

2058 AGAATGGTATGACTTTTTTATCGTGAGCCAGGCTGTGAGAACTGGTAGTCT

224 AGAATGGTATGACTTTTATCGTGAGCCAGGCTGTGAGAAGTGGTAGTGT

2118 ACATTACAATGTCATCTATGACAAACAGCGGCTGAAGCCAGACCATACAC

164 ACATTACAATGTCTATATGACAAACAGCGGCTGAAGCCAGACCACTACAC

2178 CTACA 2182

104 CIAA 100

RESULT 15

Accession	Length	Library
AF03301	640 bp	Linear
AF03301	640 bp	mRNA
AF03301	640 bp	Linear

DKFZp686G1235-11 888 (synonym: HCCS) Homo sapiens CD

REVISION
AL705301.1
AL705301.1 GI:19688656

SOURCE Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; E

REFERENCE 1 (bases 1 to 640)

Wiemann, S.
EGT (Ottenwald) Dr. Chemiker Dr. Max H. W. Weill

Published
(S.)
INTERNATIONAL JOURNAL

COMMENT
Contact: Ottenwaelder B
MTPS

Ingorslaeter Landstr.1, D-83764 Neuherberg, Germany

CRONE FROM S: WIEMANN, MOLECULAR GENOME ANALYSIS, GER
Research Center (PKFZ): Email s.wiemann@kfz- heidelberg

sequenced by MedGenomix (Munich, Germany) within the framework of the German Genome Project. N

This clone (DKFZp686G1235) is available at the RZPD i

Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de

source
1. .640

```
/mol_type="mRNA"
```


GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2003, 09:02:07 ; Search time 8509 seconds
(without alignments)
11192.574 Million cell updates/sec

Title: US-10-043-774B-1

Perfect score: 2328

Sequence: 1 atgatctttgggtgaacac.....accgcctttactactctctaa 2328

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 segs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl :

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2328	100.0	2328	9	AF264004	AF264004 Homo sapi
2	2316.2	99.5	3421	9	AF387507	AF387507 Homo sapi
3	2313	99.4	3362	9	AF104260	AF104260 Homo sapi
4	2313	99.4	3399	9	BC028581	BC028581 Homo sapi
5	2200.2	94.5	2579	9	AK093133	AK093133 Homo sapi
6	1684.6	72.4	3910	10	AB032604	AB032604 Mus muscu
7	1684.6	72.4	4064	10	AF438405	AF438405 Mus muscu
8	1013.2	43.5	3273	5	AF336369	AF336369 Danio rer
9	880	29.2	3006	9	BC031060	BC031060 Homo sapi
10	680	29.2	3138	9	AK093475	AK093475 Homo sapi
11	616.8	26.5	2456	9	AK096837	AK096837 Homo sapi
12	568.6	24.4	3057	3	AY014899	AY014899 Strongylo
13	503.2	21.6	3458	3	AK116906	AK116906 Ciona int
14	499.4	21.5	2605	3	AY014900	AY014900 Strongylo
15	460.6	19.8	4913	10	AB032605	AB032605 Mus muscu
16	439.8	18.9	3610	9	BC025995	BC025995 Homo sapi
17	435.2	18.7	3620	6	AX714218	AX714218 Sequence
18	435.2	18.7	3620	9	AK056418	AK056418 Homo sapi
19	381	16.4	3649	6	AX247684	AX247684 Sequence
20	381	16.4	3659	10	AF285586	AF285586 Mus muscu
21	379.4	16.3	1723	3	AY014901	AY014901 Strongylo
22	330.2	14.2	1455	5	BC049495	BC049495 Danio rer
23	330	14.2	2272	6	BD157952	BD157952 Primer fo
24	330	14.2	2272	9	AK027497	AK027497 Homo sapi
25	310.6	13.3	2761	3	AF145680	AF145680 Drosophil
26	257	11.0	3047	3	AF104354	AF104354 Drosophil
27	237.2	10.2	1733	6	BD156024	BD156024 Primer fo
28	237.2	10.2	1733	9	AK001213	AK001213 Homo sapi
29	222.4	9.6	63578	9	AC127071	AC127071 Homo sapi
30	220.8	9.5	193697	2	AC025837	AC025837 Homo sapi
31	203.4	8.7	61995	2	AC090565	AC090565 Homo sapi
32	184	7.9	7781	3	DMNARPL9	X94613 D.melanogas
33	184	7.9	37980	3	AC005116	AC005116 Drosophil
34	184	7.9	78507	2	AC019704	AC019704 Drosophil
35	184	7.9	169210	3	AC092226	AC092226 Drosophil
36	184	7.9	176306	3	AC007186	AC007186 Drosophil
37	184	7.9	289460	3	AE003630	AE003630 Drosophil
38	182.4	7.8	901	3	AF334408	AF334408 Drosophil
39	171.8	7.4	548	6	AF387649	AF387649 Sequence
40	170.6	7.3	186252	2	AC116715	AC116715 Mus muscu
41	170.6	7.3	211319	2	AC111089	AC111089 Mus muscu
42	164.4	7.1	259213	2	AC096050	AC096050 Rattus no
43	161.4	6.9	3677	2	AC017885	AC017885 Drosophil
44	142.2	6.1	9359	3	AF104355	AF104355 Drosophil
45	138.4	5.9	39365	2	AC100141	AC100141 Mus muscu

ALIGNMENTS

RESULT 1
AF264004
LOCUS
DEFINITION Homo sapiens HIWI mRNA, complete cds.
ACCESSION AF264004
VERSION AF264004.1 GI:15216446
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2328)
AUTHORS Sharma,A.K., Nelson,M.C., Brandt,J.E., Wessman,M., Mahmud,N.,
Weller,K.P. and Hoffman,R.
TITLE Human CD34+ stem cells express the hiwi gene, a human homologue of

the Drosophila gene piwi
Blood 97 (2), 426-434 (2001)
20584579
MEDLINE
PUBMED
11154219
2 (bases 1 to 2328)
Sharma, A.K. and Hoffman, R.
Direct Submission
Submitted (05-MAY-2000) Medicine, UIC, 900 South Ashland Avenue,
Chicago, IL 60607, USA
Location/Qualifiers
1. .2328
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="testis"
1. .2328
/codon_start=1
/product="H1WI"
/protein_id="AAK92281.1"
/db_xref="GI:15216447"
/translation="MIFGVNTRQNLHDKVHESKTSIIIVRLSTNHFLRSLTSPQWALY
QVHIDNPLMEARRLRKLSALLFOHEDLIGKHAFLPDGILFLPKRLOOKVTEVFSKTRNG
EVRITITLNELEPPTSCLOFYNIIFRLKINMLQOIGRNVXNPNPIDI PSRL
VWPGFTSILQVENSIMLCTDVSHKVLSEVLDPMFNYHQTEHKEHQEVSKELI
GLVFLTKNNKYRVDIDWDONPKSTFKADGSEVSFLRYRKQNTQELITDLQPV
VSOPKRRPGGTLPGFAMLIPELCYLTGLTDKMRNDFNMKDLAVHTLEPEQORE
VGRDITIHKNVRELNDWGLSFDNSLISFSGRILOTEKIHQSGKTFDYNPOFADW
SKETRGAPLISVPLKMLLIYTRRYEANSILQNLKFKVTPAMGMQEKALMIEVDD
RTEAVLRVQQKQVADTQIVWCLLSNNRDKYDAIKKYLCTDCTPSPQCVARTLQKQ
QTMATATKIALQNMCKMGLWRVDIPLKLVNIGVDCVHDMTAGRRSIAGFVASIN
EGMTWFSRCIQDRCQELVDGLKVLQALRAWNSCNEPMSRIIVRDGVGDGOLK
TLNVEVPQDLCKLRSIGRGNFRLTVIVKKRVNTRFFAQSGGRQLNPLPGTVIDVE
VLRPEVDFIYSQVRSVSTHYNVINDSGLKPDHIOQLTYKLCHYYNWPGEVI
RVPAPQYAKHLAFLVQSGIHRPNLSNRLYYL"
BASE COUNT 716 a 496 c 520 g 596 t
ORIGIN

Query Match 100.0%; Score 2328; DB 9; Length 2328;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2328; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGATCTTTGGTGAACACAGGAGAGACTTAGACCATCTTAAGCAATCAAAACAGGT 60
Db 1 ATGATCTTTGGTGAACACAGGAGAGACTTAGACCATCTTAAGCAATCAAAACAGGT 60
QY 61 TCTTCAGGCATTATAGTAGTTAAGCACTAACCAATTCGGGCTGACATCCCGTCCCCAG 120
Db 61 TCTTCAGGCATTATAGTAGTTAAGCACTAACCAATTCGGGCTGACATCCCGTCCCCAG 120
QY 121 TGGGCTTATATCAGTATCATTGATATACCCACTGATGAAGCGAGAGACTCCGT 180
Db 121 TGGGCTTATATCAGTATCATTGATATACCCACTGATGAAGCGAGAGACTCCGT 180
QY 181 TCAGCTCTCTTTTCAACACGAAAGTCTAATTTGGAAGTGTCTATGCTTTTGTGGAACG 240
Db 181 TCAGCTCTCTTTTCAACACGAAAGTCTAATTTGGAAGTGTCTATGCTTTTGTGGAACG 240
QY 241 ATATATTTTAACTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTGTAGTAAGCCGG 300
Db 241 ATATATTTTAACTAAAAGACTACAGCAAAAGGTTACTGAAGTTTTTGTAGTAAGCCGG 300
QY 301 AATGGAGAGATGTGAGGATTAAGATCACTTTAAACAAATGAATTCACACTACATCACCA 360
Db 301 AATGGAGAGATGTGAGGATTAAGATCACTTTAAACAAATGAATTCACACTACATCACCA 360
QY 361 ACTTGTTCAGTCTCATATATATTTTTCAGAGGCTTTTGAATCATGATTTGCAA 420
Db 361 ACTTGTTCAGTCTCATATATATTTTTCAGAGGCTTTTGAATCATGATTTGCAA 420
QY 421 CAAATTTGGAGCAATTTATATACCCCAATGACCCCAATTTGATTTTCAAGTTCAGGTTG 480
Db 421 CAAATTTGGAGCAATTTATATACCCCAATGACCCCAATTTGATTTTCAAGTTCAGGTTG 480

QY 481 GTGATTTGGCTGGCTTCACTACTTCCATCTCTAGTATGAAAAACAGCATCATGCTCTGC 540
Db 481 GTGATTTGGCTGGCTTCACTACTTCCATCTCTAGTATGAAAAACAGCATCATGCTCTGC 540
QY 541 ACTGAGCTTAGCCATAAAGTCTCTCGAAGTGAGACTGTTTGGATTTCAATGTTCAACTTT 600
Db 541 ACTGAGCTTAGCCATAAAGTCTCTCGAAGTGAGACTGTTTGGATTTCAATGTTCAACTTT 600
QY 601 TATCATCAGACAGAAACATATAATTTCAAGAAACAAGTTTCCAAAGAACTAATAGTTTA 660
Db 601 TATCATCAGACAGAAACATATAATTTCAAGAAACAAGTTTCCAAAGAACTAATAGTTTA 660
QY 661 GTTGTCTTCAAGATATAAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 720
Db 661 GTTGTCTTCAAGATATAAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 721 AATCCCAAGAGCAGCTTTAAGAAACGCGCTCTGAAAGTCAAGTCTTCTAGAAATCTAC 780
Db 721 AATCCCAAGAGCAGCTTTAAGAAACGCGCTCTGAAAGTCAAGTCTTCTAGAAATCTAC 780
QY 781 AGGAAAGCAATACAAACAGAGATCAACGACTTGAAGCAGCTCTTGTGTGAGCAGCC 840
Db 781 AGGAAAGCAATACAAACAGAGATCAACGACTTGAAGCAGCTCTTGTGTGAGCAGCC 840
QY 841 AAGAAAGCGGGGCGCTTGGGGGAGCATCTGCCAGGCGCTGCCATGCTCATTTCTGAGCTC 900
Db 841 AAGAAAGCGGGGCGCTTGGGGGAGCATCTGCCAGGCGCTGCCATGCTCATTTCTGAGCTC 900
QY 901 TGCTATCTTACAGGCTCACTAATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Db 901 TGCTATCTTACAGGCTCACTAATGATGATGATGATGATGATGATGATGATGATGATGAT 960
QY 961 GCGGTTTCATACAGACTAATCTCCAGACAAAGCAGCGTGAAGTGGGAGCTCATTTGAT 1020
Db 961 GCGGTTTCATACAGACTAATCTCCAGACAAAGCAGCGTGAAGTGGGAGCTCATTTGAT 1020
QY 1021 TACATTTCAAAAACGATTAATGTTCAAAAGGAGCTTCCAGACTGGGGTTTGGAGCTTTGAT 1080
Db 1021 TACATTTCAAAAACGATTAATGTTCAAAAGGAGCTTCCAGACTGGGGTTTGGAGCTTTGAT 1080
QY 1081 TCCAACTTTACTGCTCTCTCAGGAGAGATTTTGAACAGAGAAATTCACCAAGGTGA 1140
Db 1081 TCCAACTTTACTGCTCTCTCAGGAGAGATTTTGAACAGAGAAATTCACCAAGGTGA 1140
QY 1141 AAAACATTTGATTAACAATCCCAATTTGAGAGTGGTCCAAAGAAACAGAGGTGCACCA 1200
Db 1141 AAAACATTTGATTAACAATCCCAATTTGAGAGTGGTCCAAAGAAACAGAGGTGCACCA 1200
QY 1201 TTAATAGTGTAAAGCAGTAACTGGCTGTGATCTATACCGGAGAAATTAATGA 1260
Db 1201 TTAATAGTGTAAAGCAGTAACTGGCTGTGATCTATACCGGAGAAATTAATGA 1260
QY 1261 GAGGCAATTTCAATGATCAAAATCTATTTAAAGTTTACACGAGCATGGGCAATTAATGA 1320
Db 1261 GAGGCAATTTCAATGATCAAAATCTATTTAAAGTTTACACGAGCATGGGCAATTAATGA 1320
QY 1321 AGAAAAGCAATTAATGATTAAGTGGATGACAGAACTGAAGCTTCTTAAGAGTCTTACAG 1380
Db 1321 AGAAAAGCAATTAATGATTAAGTGGATGACAGAACTGAAGCTTCTTAAGAGTCTTACAG 1380
QY 1381 CAAAAGGTTCAGGAGACACCCAGATAGTTGTCTGTCTGTGTGTGTGTGTGTGTGTGTGTGT 1440
Db 1381 CAAAAGGTTCAGGAGACACCCAGATAGTTGTCTGTCTGTGTGTGTGTGTGTGTGTGTGT 1440
QY 1441 AAATACGATGCTTATTAATAAATACCTGTGTACAGATTCCTTACCCCAAGTCAGTGTGTG 1500
Db 1441 AAATACGATGCTTATTAATAAATACCTGTGTACAGATTCCTTACCCCAAGTCAGTGTGTG 1500
QY 1501 GTGGCCCGAACTTAAAGCAACAGCAAACTGTCTAGTGGCATTGTCTACAAAGATTCGCCCTA 1560
Db 1501 GTGGCCCGAACTTAAAGCAACAGCAAACTGTCTAGTGGCATTGTCTACAAAGATTCGCCCTA 1560
QY 1561 CAGATGAATGCAAGATGGGAGGAGAGCTTGGAGGGTGGACATCCCTCCCTGAAGTCTGTG 1620

Jiangsu 210029, China
 Location/Qualifiers
 1. .3421
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="testis"
 /dev_stage="adult"
 64. .2649
 /note="higher expression in adult testis than 6 month old
 embryo testis"
 /codon_start=1
 /product="PWI protein"
 /protein_id="AAK69348.1"
 /db_xref="GI:14579645"
 /translation="MTGRARARARGARGOETALQVGSTASQQQGYIQRPQPAPPAGV
 ELFGRCRGRTAGGTAKSOGLOISAGFQELSLAERGRRDFHDLGVNTRQNLDHVKR
 SKTSGSGLIVRLSTNHFRLISRPQWALYQYHIDYNPLMEARRLSALFOHEDLIGK
 HAPKGLPLPKLQKQKIVTSKRGVNDVRITLTNELPSTPCTLQFYNIIFRR
 LUKIMNGLQIRNYNPNDFIDIPSHRLVNPQFTTSLQYENSIMLCTDVSHKVLRS
 ETDLDFMFNFYHGTSEHKFEQVSKELIGLVLTIKYNNKTRVDDIDWDQNPSTFKK
 ADGSESVFYVRKYQNPQITLDKQPLVSKPQRRPGGTLPGPAMLIPELCVLTGK
 TDKMRNDENVMDLAVHTRLTPQORQREVGRLIDYTHKNDVORELHDDWGLSPDNLIL
 SSGRGLIQTEKHIOGSKTIDYNPQPADMSKTEGAPLSVYKPLDNLWLLIIVTRNRYEA
 NSLIQNLFPFANGQMRKALMEVDDRTREYLRVLCQQVQTDATQIVLCVLLSNRKK
 KYDAIKYICDTCPTSCQVCAVTKQQTQWMAITKALOMNCKMGSELWRVDFLQK
 LVNIVGIDYCHDGTAGRSIAGFVASINEGRTWFSRCIFQDRGEOLVDELKVCLOAK
 LRAWNSCNMPSRIIVYRDGVGDGQLKTLVNYVEVDFLCLKSIGRGNPRLTIVY
 KKEVNTRFPAQSGGRIQNPPLGTVIDVEVTRPEWYDFFIVSQAVRSGSVSPTHNYTV
 DNSGLAPDHIQRITYKLCHLYNNPQGVIRVAPQCAVAKLAFLVQGSIHREPNLSLNS
 RUYL"
 polyA signal 3354. .3359
 BASE COUNT 1071 a 661 c 767 g 922 t
 ORIGIN

Qy	188		CAACACAGAGATCTAAATCGAAGATG.CATGCTTTGATCGAAGATTTAT	241
Db	509	TTCTTTTTCACACGAGATCTAAATCGAAGTGCCATGCTTTTGTATGGAACGATATTAT	568	
Qy	248	TTTTTACCTAAAAGACTACAGCAAAAGGTTTACTGAAAGTTTTAGTAAAGCCGGAAATGGAG	307	
Db	569	TTTTTACCTAAAAGACTACAGCAAAAGGTTTACTGAAAGTTTTTGTAGTAAAGCCGGAAATGGAG	628	
Qy	308	AGGATGTGGAGTAAAGATCACCTTTTAACAANTGAACCTTCCACCTTACATCAACCACTTGTT	367	
Db	629	AGGATGTGGAGTAAAGATCACCTTTTAACAANTGAACCTTCCACCTTACATCAACCACTTGTT	688	
Qy	368	TGCAGTTCTATAATTATTATTTTCAGGAGGCTTTTGAAAATCATGAAATTTGCAACAAATG	427	
Db	689	TGCAGTTCTATAATTATTATTTTCAGGAGGCTTTTGAAAATCATGAAATTTGCAACAAATG	748	
Qy	428	GAGCAATTTATTATAACCCAAATGACCCCAATTCGATATTTCCAAAGTCACAGTTTGTGTAATTT	487	
Db	749	GAGCAATTTATTATAACCCAAATGACCCCAATTCGATATTTCCAAAGTCACAGTTTGTGTAATTT	808	
Qy	488	GGCTGGCTTCACCTACTTCATCCTTCAGTATGAAAACAGCATCATCTCTGCACTGACG	547	

Db 809 GGCTGGCTTCACTACTTCCATCCTTCAGTATGAACACAGCATCATGCTCTGCACTGAGC 868
Qy 548 TTAGCCATAAAGTCCCTTGAAGTGAGACTGTTTGGATTTCATGTTCAACTTTTATCATC 607
Db 869 TTAGCCATAAAGTCCCTTGAAGTGAGACTGTTTGGATTTCATGTTCAACTTTTATCATC 928
Qy 608 AGACAGAGAACATAAATTTCAAGACAAGATTTCACAAAGAACTAATAGGTTAGTTGTTT 667
Db 929 AGACAGAGAACATAAATTTCAAGACAAGATTTCACAAAGAACTAATAGGTTAGTTGTTT 988
Qy 668 TTACCAGATATAAATAAAGACATACAGAGTGATGATATTTGACTGGGACCAAGTCCCA 727
Db 989 TTACCAGATATAAATAAAGACATACAGAGTGATGATATTTGACTGGGACCAAGTCCCA 1048
Qy 728 AGACACCTTTAAGAAAGCCGAGGCTCTGAAGTCAGCTTCTTGAATATCTACAGGAAGC 787
Db 1049 AGACACCTTTAAGAAAGCCGAGGCTCTGAAGTCAGCTTCTTGAATATCTACAGGAAGC 1108
Qy 788 AATAACACAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCAGCCCAAGAGAA 847
Db 1109 AATAACACAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCAGCCCAAGAGAA 1168
Qy 848 GCGGGGCGCTGGGGGACACTGCCAGGGCTGCCATGCTCATTCCTGAGCTGCTGCTATC 907
Db 1169 GCGGGGCGCTGGGGGACACTGCCAGGGCTGCCATGCTCATTCCTGAGCTGCTGCTATC 1228
Qy 908 TTACAGGCTTAACCTGATAAATCGTAATGATTTTAAACGTGATGAAGACTTACGCGGTC 967
Db 1229 TTACAGGCTTAACCTGATAAATCGTAATGATTTTAAACGTGATGAAGACTTACGCGGTC 1288
Qy 968 ATACAGACTTAACCTCAGAGCAAGGAGCGTGAAGTGGGACACTCATTTGATTAATTC 1027
Db 1289 ATACAGACTTAACCTCAGAGCAAGGAGCGTGAAGTGGGACACTCATTTGATTAATTC 1348
Qy 1028 ATAAAAACGATAATGTTCAAAGGCGCTTCGAGACTGGGGTTTGGCTTTGATTTCAACT 1087
Db 1349 ATAAAAACGATAATGTTCAAAGGCGCTTCGAGACTGGGGTTTGGCTTTGATTTCAACT 1408
Qy 1088 TACTGTCCTTCTCAGAGAAATTTGCAACAGAAAGATTCACCAAGGTGGAAAAACAT 1147
Db 1409 TACTGTCCTTCTCAGAGAAATTTGCAACAGAAAGATTCACCAAGGTGGAAAAACAT 1468
Qy 1148 TTGATTACATCACAAATTTGCAAGTGGTGTGTCGCAAGAAACAGAGGTGCACCAATTA 1207
Db 1469 TTGATTACATCACAAATTTGCAAGTGGTGTGTCGCAAGAAACAGAGGTGCACCAATTA 1528
Qy 1208 GTGTTAAGCCACTAGATAACTGCTGTGATCTATACGGAAGAAATTTGAGAGCAGCCA 1267
Db 1529 GTGTTAAGCCACTAGATAACTGCTGTGATCTATACGGAAGAAATTTGAGAGCAGCCA 1588
Qy 1268 ATTCATTGATACAAATCTATTTAAAGTTACACAGCCATGGGCATGCAAAATGAGAAAA 1327
Db 1589 ATTCATTGATACAAATCTATTTAAAGTTACACAGCCATGGGCATGCAAAATGAGAAAA 1648
Qy 1328 CAATATGATTTGAAGTGGATGACAGACTGAAGCTTACTTTAGAGTCTTACAGCAAAAG 1387
Db 1649 CAATATGATTTGAAGTGGATGACAGACTGAAGCTTACTTTAGAGTCTTACAGCAAAAG 1708
Qy 1388 TCACAGCAGACACCCAGATAGTTGCTGTGTTGTCAAGTATCGGAAGGACAAATACG 1447
Db 1709 TCACAGCAGACACCCAGATAGTTGCTGTGTTGTCAAGTATCGGAAGGACAAATACG 1768
Qy 1448 ATGCTATTAAAAATACCTGTGTACAGATTGCGCTACCCCAAGTCAAGTGTGGTGGCCC 1507
Db 1769 ATGCTATTAAAAATACCTGTGTACAGATTGCGCTACCCCAAGTCAAGTGTGGTGGCCC 1828
Qy 1508 GRACCTTAGCAACAGCAAACTGTATGCGCAATGCTACAAAGATGCGCTTACAGATGA 1567
Db 1829 GRACCTTAGCAACAGCAAACTGTATGCGCAATGCTACAAAGATGCGCTTACAGATGA 1888
Qy 1568 ACTGCAAGTGGAGGAGAGCTCTGGAGGCTGACATCCCTGAGAGCTCGTATCATCG 1627
Db 1889 ACTGCAAGTGGAGGAGAGCTCTGGAGGCTGACATCCCTGAGAGCTCGTATCATCG 1948

Qy 1628 TTGGCATCGATTGTTTACCATGATCAGACTGGCGAGGTCAATCGCAGGATTTGTTG 1687
Db 1949 TTGGCATCGATTGTTTACCATGATCAGACTGGCGAGGTCAATCGCAGGATTTGTTG 2008
Qy 1688 CCAGCATCAATGAAGGATGACCCGCTGGTTCACCGTGCATATTTTCAAGATAGAGGAC 1747
Db 2009 CCAGCATCAATGAAGGATGACCCGCTGGTTCACCGTGCATATTTTCAAGATAGAGGAC 2068
Qy 1748 AGGAGCTGTAGATGGCTCAAGTCTGCTCAAGCGCTCTGAGGGCTTGAATAGCT 1807
Db 2069 AGGAGCTGTAGATGGCTCAAGTCTGCTCAAGCGCTCTGAGGGCTTGAATAGCT 2128
Qy 1808 GCAATGAGTACATGCGCAGCGGATCATGCTACCGCATGGCTAGGAGCGGCCAGC 1867
Db 2129 GCAATGAGTACATGCGCAGCGGATCATGCTACCGCATGGCTAGGAGCGGCCAGC 2188
Qy 1868 TGAAAAACACTGGTGAACCTACGAAGTGCCACAGTTTTTGGATTGTCTAAATCCATTTGTA 1927
Db 2189 TGAAAAACACTGGTGAACCTACGAAGTGCCACAGTTTTTGGATTGTCTAAATCCATTTGTA 2248
Qy 1928 GAGGTTACAAACCTAGACTTAACGGTAATTTGGTGAAGAAAGAGTGACACACAGATTTT 1987
Db 2249 GAGGTTACAAACCTAGACTTAACGGTAATTTGGTGAAGAAAGAGTGACACACAGATTTT 2308
Qy 1988 TTGCTCAGTCTGGAGAAAGACTTCAGAATCCACTTCCTGGAAACAGTTATTGATGATAGAG 2047
Db 2309 TTGCTCAGTCTGGAGAAAGACTTCAGAATCCACTTCCTGGAAACAGTTATTGATGATAGAG 2368
Qy 2048 TTACAGACACAGATGGTATGACTTTTATGCTGAGCAGCGCTGTGAGAAGTGGTAGTG 2107
Db 2369 TTACAGACACAGATGGTATGACTTTTATGCTGAGCAGCGCTGTGAGAAGTGGTAGTG 2428
Qy 2108 TTTCTCCACACATTAATGATGATCTATGACAAAGCGGCTGAAGCCAGACCATAC 2167
Db 2429 TTTCTCCACACATTAATGATGATCTATGACAAAGCGGCTGAAGCCAGACCATAC 2488
Qy 2168 AGCGTGTGACCTTACAGCTGTGCCACATCTATTACACTGGCCAGGTGTCAATCGGTTC 2227
Db 2489 AGCGTGTGACCTTACAGCTGTGCCACATCTATTACACTGGCCAGGTGTCAATCGGTTC 2548
Qy 2228 CTGCTCCTTGGCAGTACGCCACAGCTGGCTTTTCTGTTGGCCAGAGTATTACAGAG 2287
Db 2549 CTGCTCCTTGGCAGTACGCCACAGCTGGCTTTTCTGTTGGCCAGAGTATTACAGAG 2608
Qy 2288 AGCCAAATCTGTCACTGTCAACCCCTTTTACTACTCTAA 2328
Db 2609 AGCCAAATCTGTCACTGTCAACCCCTTTTACTACTCTAA 2649

RESULT 3
AF104260
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL

AF104260 3362 bp mRNA linear PRI 30-APR-2003
Homo sapiens HIWI mRNA, complete cds.
AF104260
AF104260.2 GI:18098557
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3362)
Cox,D.N., Chao,A., Baker,J., Chang,L., Qiao,D. and Lin,H.
A novel class of evolutionarily conserved genes defined by piwi are
essential for stem cell self-renewal
Genes Dev. 12 (23), 3715-3727 (1998)

2 (bases 1 to 3362)
Qiao,D., Zeeman,A.-M., Deng,W., Looijenga,L.H.J. and Lin,H.
Molecular characterization of hiwi, a human member of the piwi gene
family whose overexpression is correlated to seminomas
Oncogene 21 (25), 3988-3999 (2002)

MEADLINE 22033724
PUBMED 12037681
REFERENCE 3 (bases 1 to 3362)
AUTHORS Lin, H.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1998) Cell Biology, Duke University Medical Center, 412 Nanaline Duke Bldg., Research Dr., Durham, NC 27710, USA
REFERENCE 4 (bases 1 to 3362)
AUTHORS Qiao, D., Zeeman, A.-M., Deng, W., Looijenga, L.H.J. and Lin, H.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Cell Biology, Duke University, 412 Nanaline Duke Building, DUMC, Durham, NC 27710, USA
REMARK Nucleotide sequence updated by submitter
COMMENT On Jan 9, 2002 this sequence version replaced gi:4038412.
FEATURES
Location/Qualifiers
1..3362
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q24.33"
/tissue_type="testis"
57..2642
/note="member of the PIWI family"
/codon_start=1
/product="HIWI"
/protein_id="AAC97371.2"
/db_xref="GI:18098558"
/translation="MTGRARARARGARQOETAQLVGSTASQPGYIOPRPQPPAEG
ELFGROQRGTAGGTAKSGLOI SAGFQSLAERGRRRDFDLGVNTRQNLHVKE
STGSSGIIVRLSTNHFRITSPOWALYOHIDYNELMEARLRGALLFOHRLDGKC
HAPDTILKPKLQKQVTEVSKTENGEDVITLTNELPPTSCLOFYNIIFRR
LKTMLQOIGRNYINPNDIPISRLVINPFTSIIQYENSIMLCDSVHKVLR
ETVLPFMFYHTEHKEFOQVSKELIGLVLTINYIKYIRVDDIDWDONPKSTFK
ADGSGVFSEYRKYQNYITDLKQPLVYQPKRREGPGTLPGPAMLIPELCYLTGL
TDKMRDNFMVKDLAVHTLTPQROREVRGLDIYHKNDVQRELRDWGLSFDNLL
SPSGRLQETKHOGKTPDYNPOADWSKETRGAPLISVPLKDWLLLYTRNVEAA
NSLIQNLKFTVPMQWOMKAIMI EVDVDEAVLRVLOQVTDQTOIVVCLLSNKKO
KYDAIKYLCPTSPSCQVARTLGQQTVMALATKIALQWNCNGSBLRVDIPLK
LVNIVGIDCYHDMAGTAVGASINEGMTKRFRCIPQDRQBELVDGLKVCQAA
LWANSNCYMFPSRIIVYDVGDLGKLTIVNVEPQFLDKLSIRGRYNPLRTIVV
KRVNTRFPAQSGRLNDELPGTVIDVETREPWDFIVSQVRSQSVSPHYNYIY
DNSLKPDIHQRLTYKLCHYYNWPVIRVPAPQVAVHKLAFVIGQS IHREPLNLSN
RLYYL"
BASE COUNT 1026 a 659 c 759 g 918 t
ORIGIN
Query Match 99.4%; Score 2313; DB 9; Length 3362;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2316; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 8 TTGGTGTGAACAAGCAGAACCTTAGACCATGTTAAAGAAATCAAAACAGGTTCTTCAG 67
DB 322 TTGGTGTGAATACAAGCAGAACCTTAGACCATGTTAAAGAAATCAAAACAGGTTCTTCAG 381
QY 68 GCATTATAGTAAGGTTAAGCAGTAACCATTTCCGGCTGACATCCGTCCTCCAGTGGSCCT 127
DB 382 GCATTATAGTAAGGTTAAGCAGTAACCATTTCCGGCTGACATCCGTCCTCCAGTGGSCCT 441
QY 128 TATATCAGTATCATGATGATATACCACTGATGGAAGCAGAGACTCCGTTCCAGTCT 187
DB 442 TATATCAGTATCATGATGATATACCACTGATGGAAGCAGAGACTCCGTTCCAGTCT 501
QY 188 TTCTTTTCAACGAGAGATCTAATTGGAAGTGTGATGCTTTTGTAGTGAACCATATAT 247
DB 502 TTCTTTTCAACGAGAGATCTAATTGGAAGTGTGATGCTTTTGTAGTGAACCATATAT 561
QY 248 TTTTACTTAAGACTACAGCAAGGTTACTGAGTTTGTAGTGAACCGGAAATGGAG 307
DB 562 TTTTACTTAAGACTACAGCAAGGTTACTGAGTTTGTAGTGAACCGGAAATGGAG 621
QY 308 AGGATGTGAGGATAACGATCATTAAACAAATGAAGTCTTCCACCTTACATCAACCACTTGT 367

DB 622 AGGATGTGAGGATAACGATCATTAAACAATGAATCTCCACCTACATCAACCACTTGT 681
QY 368 TGCAGTCTTAATAATATATATTTTCAGGAGCTTTGAAAATCATGAATTTGCAACAATTTG 427
DB 682 TGCAGTCTTAATAATATATATTTTCAGGAGCTTTGAAAATCATGAATTTGCAACAATTTG 741
QY 428 GACGAAATTTATTAACCAATGACCCAAATGATATTTCAAGTCCAGGTTGGTGAATTT 487
DB 742 GACGAAATTTATTAACCAATGACCCAAATGATATTTCAAGTCCAGGTTGGTGAATTT 801
QY 488 GGCCTGGCTTCACTACTTCCATCTTCAGTATGAAAAAGCATCATGCTTCCATCACTGACG 547
DB 802 GGCCTGGCTTCACTACTTCCATCTTCAGTATGAAAAAGCATCATGCTTCCATCACTGACG 861
QY 548 TTAGCAATAAAGTCTTCGAAAGTGAAGTCTTTGGAATTCATGTTCACTTTTATCATC 607
DB 862 TTAGCAATAAAGTCTTCGAAAGTGAAGTCTTTGGAATTCATGTTCACTTTTATCATC 921
QY 608 AGACAGAAAGAACATAAAATTTCAAGAAACAAAGTTTCCAAAGAACTAATAGTTTGTGTTTC 667
DB 922 AGACAGAAAGAACATAAAATTTCAAGAAACAAAGTTTCCAAAGAACTAATAGTTTGTGTTTC 981
QY 668 TTACCAAGTATAACAATAAAGACATACAGATGATGATATTTGACCTGGGACCGAATFCCA 727
DB 982 TTACCAAGTATAACAATAAAGACATACAGATGATGATATTTGACCTGGGACCGAATFCCA 1041
QY 728 AGACACACTTTAAGAAAGCCGCGCTCTGAAGTCAAGTCTTCTTAGAATCTACAGGAGC 787
DB 1042 AGACACACTTTAAGAAAGCCGCGCTCTGAAGTCAAGTCTTCTTAGAATCTACAGGAGC 1101
QY 788 AATACAAACCAAGAGATCAACGACTTTGAAGCAGCCTGTCTTGCTCAGCAGCCCAAGAGAA 847
DB 1102 AATACAAACCAAGAGATCAACGACTTTGAAGCAGCCTGTCTTGCTCAGCAGCCCAAGAGAA 1161
QY 848 GSCGGGGCCCTGGGGGGGACATGCGCAGGCGCTGCCATGCTCTTCTGAGCTCTGCTATC 907
DB 1162 GSCGGGGCCCTGGGGGGGACATGCGCAGGCGCTGCCATGCTCTTCTGAGCTCTGCTATC 1221
QY 908 TTACAGGCTTAACCTGATTAATGCGTAAATGATTTTAAAGTGAAGAACTTAGCGGTTTC 967
DB 1222 TTACAGGCTTAACCTGATTAATGCGTAAATGATTTTAAAGTGAAGAACTTAGCGGTTTC 1281
QY 968 ATACAGACTTAACCTCAGAGCAAAAGCAGCAGCTGGAAGTGGGACGACTCATTTGATTAATTC 1027
DB 1282 ATACAGACTTAACCTCAGAGCAAAAGCAGCAGCTGGAAGTGGGACGACTCATTTGATTAATTC 1341
QY 1028 ATAAAAACGATAATGTTCAAGGGAGCTTCGAGACTGGGGTTTGAAGTTTGAATTCACACT 1087
DB 1342 ATAAAAACGATAATGTTCAAGGGAGCTTCGAGACTGGGGTTTGAAGTTTGAATTCACACT 1401
QY 1088 TACTGTCCTTCTCAGGAGAAATTTTGCBAACAGAAAGATTTCCCAAGTGGGAAACAT 1147
DB 1402 TACTGTCCTTCTCAGGAGAAATTTTGCBAACAGAAAGATTTCCCAAGTGGGAAACAT 1461
QY 1148 TTGATTAACAATCCAAATTTGAGATTTGGTCCAAAGAAACAAAGAGTGCACCATTAATTA 1207
DB 1462 TTGATTAACAATCCAAATTTGAGATTTGGTCCAAAGAAACAAAGAGTGCACCATTAATTA 1521
QY 1208 GTGTTAAGCCATAGATTAATCTGGCTTTGATCTATACCGGAAAGAAATTAAGAGAGCCA 1267
DB 1522 GTGTTAAGCCATAGATTAATCTGGCTTTGATCTATACCGGAAAGAAATTAAGAGAGCCA 1581
QY 1268 ATTCATTGATACAAATCTTAATTAAGTTTACACCGCCATGGGCATGCAATTCAGAAAG 1327
DB 1582 ATTCATTGATACAAATCTTAATTAAGTTTACACCGCCATGGGCATGCAATTCAGAAAG 1641
QY 1328 CAATATGATTCAGTGGATGACAGAACTGAAGCCCTTACTTAAGAGTCTTTACAGCAAAAG 1387
DB 1642 CAATATGATTCAGTGGATGACAGAACTGAAGCCCTTACTTAAGAGTCTTTACAGCAAAAG 1701
QY 1388 TCACAGCAGACACCCAGATAGTTGCTGCTGTCTGTTGTTCAAGTAAATCGGAAGGACCAATACG 1447

Db 399 GCATTATAGTAAGGTTAAGCAGCACTAACCAATTTCCGGCTGACATCCCGTCCCAAGTGGGCT 458
Qy 128 TATATCAGTATCAGATTGAGCTATACCACTGATGGAAGCCAGAGACTCCGTTGAGCTC 187
Db 459 TATATCAGTATCAGATTGAGCTATACCACTGATGGAAGCCAGAGACTCCGTTGAGCTC 518
Qy 188 TTCTTTTCAACAGAGATCTAAATTTGGAAGTGTGATGTTTTCATGGAAGCAATATTAT 247
Db 519 TTCTTTTCAACAGAGATCTAAATTTGGAAGTGTGATGTTTTCATGGAAGCAATATTAT 578
Qy 248 TTTTACCTAAAGACTACAGCAAAAGGTTACTGAAAGTTTATAGTAAGCCCGGAATGGAG 307
Db 579 TTTTACCTAAAGACTACAGCAAAAGGTTACTGAAAGTTTATAGTAAGCCCGGAATGGAG 638
Qy 308 AGGATGTGAGGATAAGGATCACTTTAAACAAATGAACTTCCACTACATCACCACCTTGT 367
Db 639 AGGATGTGAGGATAAGGATCACTTTAAACAAATGAACTTCCACTACATCACCACCTTGT 698
Qy 368 TGCAGTTCTATAATATTATTTTTCAGGAGCTTTTGAATAATCATGAATTTGCAACAAATG 427
Db 699 TGCAGTTCTATAATATTATTTTTCAGGAGCTTTTGAATAATCATGAATTTGCAACAAATG 758
Qy 428 GACGAAATTTATTAACCCAAATGACCCAAATGATATTCAAAGTCAACAGTTTGTGATTT 487
Db 759 GACGAAATTTATTAACCCAAATGACCCAAATGATATTCAAAGTCAACAGTTTGTGATTT 818
Qy 488 GGCCTGGCTTCACTACTTCCATCTTCAGTATGAAAAGAGCATCATGCTCTGACTGAGG 547
Db 819 GGCCTGGCTTCACTACTTCCATCTTCAGTATGAAAAGAGCATCATGCTCTGACTGAGG 878
Qy 548 TTAGCCATAAAGTCTTTGGAAGTGAAGTCTTTGGAATTCATGTTTCAACTTTTATCATC 607
Db 879 TTAGCCATAAAGTCTTTGGAAGTGAAGTCTTTGGAATTCATGTTTCAACTTTTATCATC 938
Qy 608 AGCAGAGAAGCAATAAATTTCAAGACAAAGTTTCCAAAGAACTAATAGTTTGTGTTTC 667
Db 939 AGCAGAGAAGCAATAAATTTCAAGACAAAGTTTCCAAAGAACTAATAGTTTGTGTTTC 998
Qy 668 TTACCAAGTATAACATATAGACATACAGAGTGAATGATATGATCTGGGACCAAGATCCCA 727
Db 999 TTACCAAGTATAACATATAGACATACAGAGTGAATGATATGATCTGGGACCAAGATCCCA 1058
Qy 728 AGACACCTTTAAGAAAGCCGAGCTCTGAAGTCAAGTCTTTAGAAATACTACAGGAAGC 787
Db 1059 AGACACCTTTAAGAAAGCCGAGCTCTGAAGTCAAGTCTTTAGAAATACTACAGGAAGC 1118
Qy 788 AATACAAACCAAGAGATCACCGACTTGAAGCAGCTCTGTTGGTCAGCCAGCCCAAGAGAA 847
Db 1119 AATACAAACCAAGAGATCACCGACTTGAAGCAGCTCTGTTGGTCAGCCAGCCCAAGAGAA 1178
Qy 848 GGGGGGGCTTGGGGGACACTGCCAGGGCTGCCATGCTCATCTCTGAGCTCTGCTATC 907
Db 1179 GGGGGGGCTTGGGGGACACTGCCAGGGCTGCCATGCTCATCTCTGAGCTCTGCTATC 1238
Qy 908 TTACAGGTCTAACTGATAAATGCTAAATGATTTTAACTGATAAGAACTTACCGCTTC 967
Db 1239 TTACAGGTCTAACTGATAAATGCTAAATGATTTTAACTGATAAGAACTTACCGCTTC 1298
Qy 968 ATACAGAGCTAACTCCAGAGCAAGGAGCGGTGAAGTGGGAAGCACTCATGATTAATTC 1027
Db 1299 ATACAGAGCTAACTCCAGAGCAAGGAGCGGTGAAGTGGGAAGCACTCATGATTAATTC 1358
Qy 1028 ATAAACAGATATGTTTCAAGGGAGCTTCGAGCTGGGGTTTGGCTTTGATTCACACT 1087
Db 1359 ATAAACAGATATGTTTCAAGGGAGCTTCGAGCTGGGGTTTGGCTTTGATTCACACT 1418
Qy 1088 TACTGCTCTTCTCAGGAAGAAATTTTGAACAGAAAGATTTTCAAGAGTGGGAAACAT 1147
Db 1419 TACTGCTCTTCTCAGGAAGAAATTTTGAACAGAAAGATTTTCAAGAGTGGGAAACAT 1478
Qy 1148 TTGATTACAATCCCAATTTTGCAGATTTGTTCCAAAGAAACCAAGAGTGGCAATTAATTA 1207
Db 1479 TTGATTACAATCCCAATTTTGCAGATTTGTTCCAAAGAAACCAAGAGTGGCAATTAATTA 1538

Qy 1208 GTGTTAAAGCCACTAGATAAAGCTGGCTGTGATCTATACGCGAAGAAATTTATGAGCAGCCA 1267
Db 1539 GTGTTAATCCACTAGATAAAGCTGGCTGTGATCTATACGCGAAGAAATTTATGAGCAGCCA 1598
Qy 1268 ATTCAATTGATACAAAATCTATTTTAAAGTTACACAGCCATGGGATGCAAAATGAGAAAAG 1327
Db 1599 ATTCAATTGATACAAAATCTATTTTAAAGTTACACAGCCATGGGATGCAAAATGAGAAAAG 1658
Qy 1328 CAATAATGATTGAAGTGGATGACAGAACTGAGCTTAAAGTCTTACAGCAAAAGG 1387
Db 1659 CAATAATGATTGAAGTGGATGACAGAACTGAGCTTAAAGTCTTACAGCAAAAGG 1718
Qy 1388 TCACAGCAGACACACAGATAGTCTGCTGTCTGTTGTCAAGTAAATCGGAAGGACAAATACG 1447
Db 1719 TCACAGCAGACACACAGATAGTCTGCTGTCTGTTGTCAAGTAAATCGGAAGGACAAATACG 1778
Qy 1448 ATGCTATTAAAAATACTGTTGATCAGATTCCTTACCCCAAGTCACTGTGTGGTGGGCC 1507
Db 1779 ATGCTATTAAAAATACTGTTGATCAGATTCCTTACCCCAAGTCACTGTGTGGTGGGCC 1838
Qy 1508 GAACTTTAGGCAAAACAGCAAACTGTCATGGCCATTTGCTACAAAGATTCCTTACAGATGA 1567
Db 1839 GAACTTTAGGCAAAACAGCAAACTGTCATGGCCATTTGCTACAAAGATTCCTTACAGATGA 1898
Qy 1568 ACTGCAAGATGGGAGGAGAGCTCTGGAGGTTGGACATCCCTCTGAAGCTCGTGATGATCG 1627
Db 1899 ACTGCAAGATGGGAGGAGAGCTCTGGAGGTTGGACATCCCTCTGAAGCTCGTGATGATCG 1958
Qy 1628 TTGGCATCGATTTGTTACATGATGACAGCTGGGCGGAGGTTCAATCGCAGGATTTGTTG 1687
Db 1959 TTGGCATCGATTTGTTACATGATGACAGCTGGGCGGAGGTTCAATCGCAGGATTTGTTG 2018
Qy 1688 CCAGCATCAATGAAGGATGACCCGCTGCTCTCAAGCGGCTCTGAGGGCTTGGATGATGCT 1747
Db 2019 CCAGCATCAATGAAGGATGACCCGCTGCTCTCAAGCGGCTCTGAGGGCTTGGATGATGCT 2078
Qy 1748 AGGAGCTGTTAGTGGGCTCAAGTCTGCTCAAGCGGCTCTGAGGGCTTGGATGATGCT 1807
Db 2079 AGGAGCTGTTAGTGGGCTCAAGTCTGCTCAAGCGGCTCTGAGGGCTTGGATGATGCT 2138
Qy 1808 GCAATGAGTACATGCCAGCCGAGTCACTGTTGATCCGAGTGGGCTAGGAGCGGCCAGC 1867
Db 2139 GCAATGAGTACATGCCAGCCGAGTCACTGTTGATCCGAGTGGGCTAGGAGCGGCCAGC 2198
Qy 1868 TGAAAACACTGTTGAACTTACGAAAGTCCACAGTCTTGGATTTGCTAAATCCATTTGTTA 1927
Db 2199 TGAAAACACTGTTGAACTTACGAAAGTCCACAGTCTTGGATTTGCTAAATCCATTTGTTA 2258
Qy 1928 GAGGTTACAAACCTTAGACTTAAACGTTAATTTGGTGAAGAAAAGAGTGAACACAGATTTT 1987
Db 2259 GAGGTTACAAACCTTAGACTTAAACGTTAATTTGGTGAAGAAAAGAGTGAACACAGATTTT 2318
Qy 1988 TTGCTCAGTCTGGAGGAAGACTTTCAGATTCACCTTCTGGAACAGTTATGATGATAGAG 2047
Db 2319 TTGCTCAGTCTGGAGGAAGACTTTCAGATTCACCTTCTGGAACAGTTATGATGATAGAG 2378
Qy 2048 TTACAGACCCAGAAATGGTATGACTTTTATCGTGAGCCAGGCTGTGAAAGTGGTAGTG 2107
Db 2379 TTACAGACCCAGAAATGGTATGACTTTTATCGTGAGCCAGGCTGTGAAAGTGGTAGTG 2438
Qy 2108 TTTTCTCCCAACATTAATGTCATCTATGACAAAGCGGCTTGAAGCCAGACCAATATAC 2167
Db 2439 TTTTCTCCCAACATTAATGTCATCTATGACAAAGCGGCTTGAAGCCAGACCAATATAC 2498
Qy 2168 AGCGCTTGAACCTTACAGGCTGTCGCACTATTAACATGCGCAAGTGTCAATTCGTGTTTC 2227
Db 2499 AGCGCTTGAACCTTACAGGCTGTCGCACTATTAACATGCGCAAGTGTCAATTCGTGTTTC 2558
Qy 2228 CTGCTCTTGGCCAGTACGCCCAAGCTGGCTTTTCTTGTGGCCAGAGTATTCAGAG 2287
Db 2559 CTGCTCTTGGCCAGTACGCCCAAGCTGGCTTTTCTTGTGGCCAGAGTATTCAGAG 2618

QY 2288 AGCCBACCTCTGCTCAACCGGCTTTACTACCTCTAA 2328
 Db 2619 AGCCAAATCTGCTCAACCGGCTTTACTACCTCTAA 2659

RESULT 5
 AK093133
 LOCUS Homo sapiens cDNA FLJ35814 fis, clone TESTI2006083, highly similar
 to Mus musculus mRNA for MIWI (piwi).

ACCESSION AK093133
 VERSION AK093133.1 GI:21751907
 KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Mammalia; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
 AUTHORS Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,
 Arita, M., Musashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R.,
 Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,
 Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,
 Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
 Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B.,
 Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and
 Isogai, T.

TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2579)
 AUTHORS Isogai, T. and Yamamoto, J.
 DIRECT SUBMISSION Direct Submission

TITLE Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
 JOURNAL Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
 COMMENT (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology (RAB); cDNA library
 construction: Helix Research Institute (HRI) (supported by Japan
 Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
 HRI, and Biotechnology Center, National Institute of Technology and
 Evaluation; clone selection for full insert sequencing: HRI and
 RAB; annotation: HRI and RAB.

FEATURES
 Location/Qualifiers
 1..2579
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="TESTI2006083"
 /tissue_type="testis"
 /clone_lib="TESTI2"
 /notes="Cloning vector: pME18SFL3"
 54..2543
 /notes="unnamed protein product"
 /codon_start=1
 /protein_id="BAC04068.1"
 /db_xref="GI:21751908"
 /translation="MTGRARARARGRGQETALVNGSTASQOPVHQPQPAPBAG
 ELFGKRGQTAGTAKSLQISAGFQELSLAERGRRRDFDLGVNTRQLNDHVKC
 SKTSSGIIIVRLTNHFRSPQWALYQYHIDYNPLMEARLRSLALFQHEDLIKC
 HAPDGLILFLPKLQKQKATEVFSKTRNGEDVRITITLNEPLPPTSPTCLQFYNIIPRR
 LKIMLQOIGRNYVNPDPIDIPSHRLVTPGFTTSILQYENSIMLCTDVSHKVLRS
 ETVLDFWENFYHCTEHKQEQVSKELIGLVLTIKNNKYRYRVDLDWDQNPSPFKK
 ADGVSFVLEIYRKQYNEITLDEKQVLSVQPKRRPGGTLPGPAMLIPELCYLTGL
 TDQRNDFNMDLAVHTLITPEQREVRGLIDYTHKNNDVQRELDWGLSFDNSLL
 SFSGRITQEKIHKQKTDYNDPQADWSKETRGAPLISVKPLDNLMLLYTRRNYEA
 NSLIQNTQKTPAMQMKRAIMIEVDVTEALRVLOQKVADTQIVVCLLSNKKD
 KYDAIKYLCDCPTQSQVAFELCKQCTVMAIAIKALQWCKXGSELNWDIPLK
 LWNIVGIDCHWAGRSSTAGFVASINEGMTNFRSCIFQDRGQSLDGLKVCLOQA
 LRWNSCNEMPRRIIVIRVDGVDGQKTLVNYEVPQFLDCLKISGRGNPRUTIV
 KGVNTRFPAQSGRLQNLQEPGTVIDVETVRPEWYDFIVSQAVRSQSVSPHYNYI
 DNSGLKPDHITQRIYKLCIHYNNPVSASTC"

CDS
 54..2543
 /notes="unnamed protein product"
 /codon_start=1
 /protein_id="BAC04068.1"
 /db_xref="GI:21751908"

BASE COUNT 797 a 542 c 613 g 627 t
 ORIGIN

Query Match 94.5%; Score 2200.2; DB 9; Length 2579;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2202; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 8 TTGGTGTGACACAGGCGAAGACTTTAGACCATGTGTTAAAGAAATCAAAGAGTCTTCTAG 67
 Db 319 TTGGTGTGATTAACAGGCGAAGACCTAGACCATGTGTTAAAGAAATCAAAGAGTCTTCTAG 378
 QY 68 GCATTATAGTAGGTAAAGCACTAACCTATTCGGGTGACATCCCGTCCCGAGTGGCCCT 127
 Db 379 GCATTATAGTAGGTAAAGCACTAACCTATTCGGGTGACATCCCGTCCCGAGTGGCCCT 438
 QY 128 TATATCAGTATCACATTCATATATACCCATGATGAAGCCAGAGACTCCGTTAGCTC 187
 Db 439 TATATCAGTATCACATTCATATATACCCATGATGAAGCCAGAGACTCCGTTAGCTC 498
 QY 188 TTCTTTTCAACACGAGAGATCTAAATGGAAAGTGTGATGCTTTTGTGAAACGATATTAT 247
 Db 499 TTCTTTTCAACACGAGAGATCTAAATGGAAAGTGTGATGCTTTTGTGAAACGATATTAT 558
 QY 248 TTTTACCTTAAAGAGACTACAGCAAAAAGGTACTGAAAGTCTTTAGTAAGACCCGGAATGGAG 307
 Db 559 TTTTACCTTAAAGAGACTACAGCAAAAAGGTACTGAAAGTCTTTAGTAAGACCCGGAATGGAG 618
 QY 308 AGGATGTGAGGATAACGATCACTTTTAAACAATGAACTTCCACCTACATCAACAATGTT 367
 Db 619 AGGATGTGAGGATAACGATCACTTTTAAACAATGAACTTCCACCTACATCAACAATGTT 678
 QY 368 TGCAGTCTTATAATATATTTTCAGAGGCTTTTCAAGAAATCATGAATTTGCAACAATG 427
 Db 679 TGCAGTCTTATAATATATTTTCAGAGGCTTTTCAAGAAATCATGAATTTGCAACAATG 738
 QY 428 GACGAAATTTATATATACCAATGACCAATGATTTTCCAAAGTCAAGTTGGTGATTT 487
 Db 739 GACGAAATTTATATATACCAATGACCAATGATTTTCCAAAGTCAAGTTGGTGATTT 798
 QY 488 GGCCTGGCTTCACTACTTCCATTCCTTCAAGTATGAAACAGCATCATGCTCTGCACTGACG 547
 Db 799 GGCCTGGCTTCACTACTTCCATTCCTTCAAGTATGAAACAGCATCATGCTCTGCACTGACG 858
 QY 548 TTAGCCATAAAGTCTTCAAGTGTGAGACTGTTTGGATTTTCATGTTTCACTTTTATCATC 607
 Db 859 TTAGCCATAAAGTCTTCAAGTGTGAGACTGTTTGGATTTTCATGTTTCACTTTTATCATC 918
 QY 608 AGACAGAGAACATATAATTTCAAGAACCAAGTTTCCAAAGAACTAATAGTGTAGTTGTTTC 667
 Db 919 AGACAGAGAACATATAATTTCAAGAACCAAGTTTCCAAAGAACTAATAGTGTAGTTGTTTC 978
 QY 668 TTACCAAGTATTAACAATAAGACATACAGAGTGATGATTTGACTGGGACCGAATCCCA 727
 Db 979 TTACCAAGTATTAACAATAAGACATACAGAGTGATGATTTGACTGGGACCGAATCCCA 1038
 QY 728 AGAGCACCTTTAAGAAAGCCGACGCTCTGAAGTCAAGCTTCTTGAATACTACAGGAAGC 787
 Db 1039 AGAGCACCTTTAAGAAAGCCGACGCTCTGAAGTCAAGCTTCTTGAATACTACAGGAAGC 1098
 QY 788 AATAACAACCAAGAGATCAACCGCTTTGAAGCAGCCTGTCTTGGTCAAGCCACCCCAAGAGAA 847
 Db 1099 AATAACAACCAAGAGATCAACCGCTTTGAAGCAGCCTGTCTTGGTCAAGCCACCCCAAGAGAA 1158
 QY 848 GCGGGGGCCCTGGGGGACACATGCGAGGCGCTGCCATGCTCTTCTGAGCTCTGCTATC 907
 Db 1159 GCGGGGGCCCTGGGGGACACATGCGAGGCGCTGCCATGCTCTTCTGAGCTCTGCTATC 1218
 QY 908 TTACAGGCTCTAATCATATAATGCGTAATGATTTTAAAGTGTGATGAAGACTTAGCGGTC 967
 Db 1219 TTACAGGCTCTAATCATATAATGCGTAATGATTTTAAAGTGTGATGAAGACTTAGCGGTC 1278
 QY 968 ATACAAGACTTAATCTCAGAGCAAAAGCGCGTGAAGTGGAGCACTCATTTGATTATCATTC 1027
 Db 1279 ATACAAGACTTAATCTCAGAGCAAAAGCGCGTGAAGTGGAGCACTCATTTGATTATCATTC 1338

Qy	1028	ATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCAACT	1087
Db	1339	ATAAAAACGATAATGTTCAAAGGGAGCTTCGAGACTGGGGTTTGAGCTTTGATTCAACT	1398
Qy	1088	TACTGTCCCTCTCAGGAGAAATTTTGCAPACAGAAAGATTTCACCAAGGTGGAAACAT	1147
Db	1399	TACTGTCCCTCTCAGGAGAAATTTTGCAPACAGAAAGATTTCACCAAGGTGGAAACAT	1458
Qy	1148	TTGATTACAATCCACAATTTGCGATTGGTCCAAAGAAAACAGAGGTGCACCAATTAATTA	1207
Db	1459	TTGATTACAATCCACAATTTGCGATTGGTCCAAAGAAAACAGAGGTGCACCAATTAATTA	1518
Qy	1208	GTGTTAAGCCATPAGATAAATCGGCTGTGTGATCTATACGCGAAGAAATTAATGAAGCGCCA	1267
Db	1519	GTGTTAAGCCATPAGATAAATCGGCTGTGTGATCTATACGCGAAGAAATTAATGAAGCGCCA	1578
Qy	1268	ATTCAATTGATACAAAATCTATTTAAAGTTACACCGCCATGGGCATGCCAAATCAGAAAG	1327
Db	1579	ATTCAATTGATACAAAATCTATTTAAAGTTACACCGCCATGGGCATGCCAAATCAGAAAG	1638
Qy	1328	CAATAATGATTGAAGTGGATGACAGAACTCGAAGCCTACTTTAAGAGTCTTACAGCAAAAGG	1387
Db	1639	CAATAATGATTGAAGTGGATGACAGAACTCGAAGCCTACTTTAAGAGTCTTACAGCAAAAGG	1698
Qy	1388	TCACAGCAGACACCCAGATAGTTGTCGTCTGTTGTCAAATAATCGGAAGGACAAATACG	1447
Db	1699	TCACAGCAGACACCCAGATAGTTGTCGTCTGTTGTCAAATAATCGGAAGGACAAATACG	1758
Qy	1448	ATGCTATTAAAAATACCTGTGTACAGATTGCCTTACCCCAAGTCAAGTGTGGTGGCCCC	1507
Db	1759	ATGCTATTAAAAATACCTGTGTACAGATTGCCTTACCCCAAGTCAAGTGTGGTGGCCCC	1818
Qy	1508	GAACCTTAGGCAACACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGA	1567
Db	1819	GAACCTTAGGCAACACAGCAAACTGTCATGGCCATTGCTACAAAGATTGCCCTACAGATGA	1878
Qy	1568	ACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGAATGATG	1627
Db	1879	ACTGCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCCCCTGAAGCTCGTGAATGATG	1938
Qy	1628	TTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTG	1687
Db	1939	TTGGCATCGATTGTTACCATGACATGACAGCTGGGCGGAGGTCAATCGCAGGATTTGTTG	1998
Qy	1688	CCAGCATCAATGAAGGATGACCCGCTGGTTCTCACGCTGCATATTTTCAGATAGAGGAC	1747
Db	1999	CCAGCATCAATGAAGGATGACCCGCTGGTTCTCACGCTGCATATTTTCAGATAGAGGAC	2058
Qy	1748	AGGAGCTGGTAGATGGGCTCAAACTGTCCCTGCAAGCGGCTCTGAGGGCTTGGAAATAGCT	1807
Db	2059	AGGAGCTGGTAGATGGGCTCAAACTGTCCCTGCAAGCGGCTCTGAGGGCTTGGAAATAGCT	2118
Qy	1808	GCAATGAGTACATGCCAGCCGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCAGC	1867
Db	2119	GCAATGAGTACATGCCAGCCGATCATCGTGTACCGCGATGGCGTAGGAGACGGCCAGC	2178
Qy	1868	TGAAAAACACTGGTGAATACGAACTGCCAGAGTGCACAGTTTTTTGGATTGTCTAAAAATCCATTGGTA	1927
Db	2179	TGAAAAACACTGGTGAATACGAACTGCCAGAGTGCACAGTTTTTTGGATTGTCTAAAAATCCATTGGTA	2238
Qy	1928	GAGGTTACAACTTACGTTAACTGTTGGTGAAGAAAAGAGTAGGAAACACCAAGATTTT	1987
Db	2239	GAGGTTACAACTTACGTTAACTGTTGGTGAAGAAAAGAGTAGGAAACACCAAGATTTT	2298
Qy	1988	TTGCTCAGTCTGGAGGAAGACTTCAGAAATCCACTTCTCTGGAAACAGTTATTGATGTAGAGG	2047
Db	2299	TTGCTCAGTCTGGAGGAAGACTTCAGAAATCCACTTCTCTGGAAACAGTTATTGATGTAGAGG	2358
Qy	2048	TTACCGAGCAGAAATGGTATGCACTTTTTTATCGTGTAGCCAGGCTGTGAGAAAGTGTGATG	2107
Db	2359	TTACCGAGCAGAAATGGTATGCACTTTTTTATCGTGTAGCCAGGCTGTGAGAAAGTGTGATG	2418
Qy	2108	TTTTCTCCACACATTACAATGTCTATGTACAAACAGCGGCTCTGAAGCAGACACCACTAC	2167

Db 2419 TTTCTCCACACATTACATGTCTATCTATGACACAGCGCTTGAGCCAGACCATAC 2478
 Qy 2168 AGCGCTTGACCTCAAGCTGTGCCACATCTATTACAACTGGCCAG 2212
 Db 2479 AGCGCTTGACCTCAAGCTGTGCCACATCTATTACAACTGGCCAG 2523

RESULT 6
 AB032604
 LOCUS AB032604 3910 bp mRNA linear ROD 04-OCT-2001
 DEFINITION Mus musculus mRNA for MIWI, complete cds.
 ACCESSION AB032604
 VERSION AB032604.1 GI:7416110
 KEYWORDS MIWI (piwi).
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Kuramochi-Miyagawa, S., Kimura, T., Yomogida, K., Kuroiwa, A.,
 Tadokoro, Y., Fujita, Y., Sato, M., Matsuda, Y. and Nakano, T.
 Two mouse piwi-related genes: miwi and miwi
 Mech. Dev. 108 (1-2), 121-133 (2001)
 21463379
 11578866
 2. (bases 1 to 3910)
 Miyagawa, S.K. and Nakano, T.
 Direct Submission
 Submitted (21-SEP-1999) Satomi K Miyagawa, Research Institute for
 Microbial Diseases, Department of Molecular Cell Biology; Yamadaoka
 3-1, Suita, Osaka 565-0871, Japan
 (E-mail:smiyagawa@biken.osaka-u.ac.jp, Tel:81-6-6879-8362,
 Fax:81-6-6879-8362)
 FEATURES
 Location/Qualifiers
 1. 3910
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /map="5F distal-G2 proximal"
 /tissue_lib="testis"
 1. 3910
 /gene="miwi"
 156..2744
 /gene="miwi"
 /note="piwi family"
 /codon_start=1
 /product="MIWI"
 /protein_id="BAA93705.1"
 /db_xref="GI:7416111"
 /translation="MTGEPARARGARGQETVHVCAABSQPGYIPRPPQSPPTSG
 DLVGRQRGVNVGATSKSSELOISAGFQBLSLAERGRRDRHDGLVNTRONLDHVNK
 ESKTSSGIIIVLSTNNHFRSTPQWALYQTHIDNPELMARRLRSLALLQHEDLIGR
 CHAFDITLFLKRLQHKVTEVSOTRNGEVRHRTITLTNELPPTSTCLOFYNIIFR
 RLAKIMNLQYIGNYNPSPDIIPNHRLVIPGFTTILQYENNIMLCITDVSHKFLP
 KATSDVFMNLVYQOTBEQFOESKELIGLIVLTNNKTYRDDIDWDPKSTFK
 SEGVSEVFLVRYKQNOEITLKKQPLVLSQPKRRGPGGTLPGPAMLLPELCYLNG
 LTKMRDNFNMKDLAVHTLTLTEQOREVREGLIDYIHKDDNVQRELRDGLGSPDSNL
 LNSFLRQISEKHQHGQETDINPQADWSKTEGAPLISVLPKLDNMLLIYTRNNSE
 ANSLQNLFFKVTGQMKAMLEIYDDTEAYLRALQOKVTSDTQIVVCLLSNNRK
 DKYDALIKYICDTCPTSPQAVARTLQKQTVMAITKALQWNCKMGGLRWVDMPL
 KLAMIVGIDCHDTTARRRISAGFVASINEGMTRFWSRQVFDGRGOELVDGLKVCLOA
 ALRAWGSCNEVMSRVIVYDVGVDGLKTLVNVVEPQFLDCLKSVGRGVNPRLTIV
 VKRVNARFFAQGGRLQNLDPGVTDIVETREPEWDFFIVLSQAVSGSVSPHYNIV
 YDSGSLKPDHQRILTKYLCHVYVNPWPAPQCYAHKLAFLVGSQSIHREPNLSLUS
 NRIYYL"

BASE COUNT 993 a 970 c 1063 g 884 t
 ORIGIN

Query Match 72.4%; Score 1684.6; DB 10; Length 3910;
 Best Local Similarity 82.8%; Pred. No. 0;
 Matches 1924; Conservative 0; Mismatches 399; Indels 0; Gaps 0;


```
Db 2582 CCAGCGGCTGACATCAAGCTCTGCCACGTGTACTATATTTGGCTGGAGTCATCCGAGT 2641
Qy 2226 TCCTGCTCTTCCGAGTACGCCCAAGCTGGCTTTCTTTGCTGGCAGAGTATTCACAG 2285
Db 2642 CCCTGCACCTTCCGAGTATGACACAAGCTGGCTTCTCGTGGGCGAGAGATCCACAG 2701
Qy 2286 AGAGCCAAATCTGTACTCTCAAAACCGCTTTACTACCTCTAA 2328
Db 2702 AGAGCCAAACCTCTCCCTGTCCAAACCGCTTACTACTACTTAA 2744

RESULT 7
LOCUS AF438405 4064 bp mRNA linear ROD 13-NOV-2001
DEFINITION Mus musculus MIWI mRNA, complete cds.
ACCESSION AF438405
VERSION AF438405.1 GI:16905060
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS Deng, W. and Lin, H.
TITLE Miwi, a murine homolog of piwi, encodes a cytoplasmic protein
essential for spermatogenesis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4064)
AUTHORS Deng, W. and Lin, H.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2001) Department of Cell Biology, Duke
University, Room 412, Nanaline Duke Building, Durham, NC 27710, USA

FEATURES
source
    /organism="Mus musculus"
    /mol_type="mRNA"
    /db_xref="taxon:10090"
    /chromosome="5"
    /tissue_type="testis"
    /note="Piwi-like protein"
    /codon_start=1
    /product="MIWI"
    /protein_id="AA131014.1"
    /db_xref="GI:16905061"
    /translation="MTGRARARARGARQETVQHVAAASQPGYIPRPOOSPTEG
    DIVGRGRGMVVGATSKSOELQISAGFQELSLAERGRDRPDLGVNTRQNLDBVK
    ESKTSSGSIIVKLTSTNHFRTSRPQWALYQYHIDVNPMLMEARRLSALLFQHEDLIGR
    CHAFDGTILFLPKRLQHKVTEVPSOTRNGEHVRIITITLNELPPTSPCLQFYNIIFR
    RLKTMNLOIGRNVNPSDPIDIPNHLVIVPGFTTSILOYENIMLCTDYSHKVLIR
    SETVLDFMNLVQTEEHKFORQVSKELGLVLTKYNNKTYRVDIDWDQNPSTFK
    KADGEVSELEYRKQINORIIDLQPVLSQPKRRRGGHLPFGPMLIPBELCLITG
    LTKMRNDFNVNKDLAVHTRLPPEQRREVRLIDYIHKDNDVQRELRLDWGLSPDSNL
    LSFSGRLQSEKHQGGKTFDYNQPADWSKETRGAPLISKVPLDNWLLIYTRRYEA
    ANSLIOLKFKVTPMGKIMIEVDDRTEAYLRALQOKVTSDTQIVVCLISNRK
    DKYDALTKFLVCTDPTPSOCVARTLGKQOTVMAJATKIALQMNCKMGGELWRVMDPL
    KIAMVIGIDYHDITAGRSIAGFVASINEGTRPSRCVDFDRGSELVDGLKVLQQA
    VUKRVNARFPAQSGRLQNPFGTIDVETPEWDFVFIQAVRSVGSVSTHNYVI
    YRSLGLKPDHIQLRLATYKLCVNYNPGVIRVPAPCOYAHKLAFLVQSGTHREPNLSLS
    NRLYL"

BASE COUNT 1115 a 982 c 1077 g 890 t
ORIGIN
Query Match 72.4%; Score 1684.6; DB 10; Length 4064;
Best Local Similarity 82.8%; Pred. No. 0;
Matches 1924; Conservative 0; Mismatches 399; Indels 0; Gaps 0;
Qy 6 CTTTGTGTGAACACAGGAGAGACTTAGACCATGTATAAGATCAAAAACAGGTTCTTC 65
Db 457 CTTTGTGTGAACACAGGAGAGACTTAGACCATGTATAAGATCAAAAACAGGTTCTTC 65
Qy 66 AGSCATTATAGTAAAGTAAAGCACTAACCATTTCCGGCTGACATCCCGTCCCGAGTGGG 125
```


Db 2539 CCAGCTCCATGCGAGTATGCCACAAATTTGGCTTCTCTGGTGGCCAGAGCATTCACAG 2598

QY 2287 GAGCCAAATCTGTCACTGTCAACCCCTTTACTACCTCTAA 2328

Db 2599 GAGCCTAACATGAATCTGGTGAATCTCTGTATTACCTGTAA 2640

BC031060 3006 bp mRNA linear PRI 13-JUN-2002

Homo sapiens, Similar to piwi like homolog 1 (Drosophila), clone

MGC:33296 IMAGE:5271278, mRNA, complete cds.

BC031060

BC031060.1 GI:21410560

MGC.

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 3006)

Strausberg,R.

Direct Submission

Submitted (03-JUN-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabbs@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki

Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,

R. M.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 48 Row: b Column: 15

This clone was selected for full length sequencing because it

passed the following selection criteria: Similarity but not

identity to protein.

Location/Qualifiers

1. .3006

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="MGC:33296 IMAGE:5271278"

/tissue_type="Testis"

/clone_lib="NIH MGC 97"

/lab_host="DH10B"

/note="Vector: pBluescript"

4. .2562

/codon_start=1

/product="Similar to piwi like homolog 1 (Drosophila)"

/protein_id="AAH31060.1"

/db_xref="GI:21410561"

/translation="MSGRARVARGIARSPSATEVGRIOASPLPSVDLSNNEASSSN

GLIPSLRSTNKGIGSSDAGSTFMERGVKNQDPMDSICTREKLAHVRNCKTGS

GIPVLVTLNFDLPDDPOLYOYHYIYIPDLASRLRIALLYSHLSNKAAPDGA

ILFLSKLEKYEYTESSETGETTITMTTLKRELSPSPVCIQVFNIIIFRILKLS

MTQIGRNFYNSPEMIPQHLKSLMFPFAISYFERKLFPADSVYKVLNRTVLEF

MTALCORTGLSCFTQCEKQJGLIIVLTRYNNRTSIDIDWSVKPHTFOKRDTEI

TVDYVYKQYDITVSDLPMLVSLKKRDNSEALAHLPCLFELTGLTDQATSD

FQIMKAVAKTLSPSGRQRLARLVNIRQNTNARFELTWGLHFGSQTSLGRIVP

SEKILMQDHCIPVSAADWSKDIRCTKILNAGSLNLTWLICSDRTEYVAFSENLRL

VAGSMGVNDYDIKIQVENPAAFVRAIQYVDPDQVLMVLCIPSNQKTYDSIKRL

SSDCPVPSOCVLARTLNKGWMSIATKIAMQNTCKLGGELWALBPLKSLMVGVIDV

CKDALSQVWVGVASVNPRTIRWFSRCILQRTMTDVADECLKVPMTGALNKWYKXNH

FEATURES

source

CDS

Query Match 29.2%; Score 680; DB 9; Length 3006;

Best local Similarity 56.7%; Fred. No. 3.7e-172;

Matches 1319; Conservative 0; Mismatches 1000; Indels 9; Gaps 3;

BASE COUNT 904 a 627 c 674 g 801 t

ORIGIN

1 ATGATCTTTTGTGTGACACACAGGAGCAATTTAGACCAATGTTAAAGAATCAAAAACAGT 60

244 ATGGATTTCAGTAATCTGTACACAGAAATAATGGCAATGTGAGAAATGTTAAACAGT 303

61 TCTTCAGGCATTTAGTAAGGTTAAGCACTAACCAATTTCCGGCTGACATCCCGTCCCCAG 120

304 TCCAGTGAAATACCTGTGAAACTGGTTACAAACCTCTTTAACTTAGATTTTCCCCAAGAC 363

121 TGGGCTTTATATCAGTATCAGATTCAGTATTAACCACTGATGAAGCAGCAGATCCGT 180

364 TGGCAGCTATACCACTACCATGTGACATATATTCAGATTTTAGCATCTAGAAAGGCTGAGA 423

181 TCAGCTCTCTTTTCAACACAGAGATCTAATTTGAAAGTGTCTATGCTTTTGTATGAGAACG 240

424 ATTGCTTTTACCTTTTACCTGATAGTGAATTTCCCAACAAGCAAGCAATTCGACGGTGC 483

241 ATATTATTTTACCTTAAAGACTACAGCAAAAGGTTACTGAAGTTTTTAGTAAGACCCGG 300

484 ATCTCTTTTCTGTACAAAGCTAGAGAAAGGTCACAGAGTTGTCAAGTGAATCAAACTCAA 543

301 AATGAGAGGATGTGAGGATAACGATCTCTTAACAATGAATCTCCACCTACATACCA 360

544 AGAGTGAGACTATAAAGATGACTATCACCTGAAGAGGGAGCTGCCATCAAGTTCTCCC 603

361 ACTGTTTGGCAGTTCTATAATATTATTTTTCAGGAGGCTTTTGAAATCATGAATTTGCAA 420

604 GTGTCATCCAGGCTTCAATATCATCTTCAGAAAGATCTCCAAAGAGTTGTCCATGTAC 663

421 CAATTCGACGAAATTTATTAACCAATGACCAATTTGATTTTCCAAAGTCACAGGTTG 480

664 CAATTCGACGAACTTCTATAATCTTTCAGAGCAATGGAATTTCCCGACGACAAATTA 723

481 GTGATTTGGCTTGGCTTCACTACTTCCATCTCTCAGTATGAAGAACAGCATCATGTCTGC 540

724 TCCCTTTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTTGGCTT 783

541 ACTGACGTTAGCATATAAAGTCTTCGAAAGTCAGACTGTTTGGATTTTCATGTTCAACTTT 600

784 GCTGATGTGAGTTACAAAGTCTTCGGGATGAGAGGTTCTTGGATTTCATGACTCTCTC 843

601 TATCATCAGACAGAAAGACATAAATTTCAAGAACAAAGTTTCCAAAGAACTAATAGGTTTA 660

844 TGTCAAGAACTGGCTTGTCTCTCTTCCACCGAGCTGTGAGAGCAGCTAATAGGGCTC 903

661 GTTGTCTTCAAGTATAAACAATAAGACATACAGAGTGGATGATTTGACTTGGGACCCAG 720

904 ATTGCTCTTCAAGATACATAAGCAACCTACTCTCAATGATGACATTTGATGGTCAGTG 963

721 AATCCAAAGAGCACCTTTTAAAGAAAGCCGAGCTCTGAAGTCAGCTTCTTAGAATACTAC 780

964 AAGCCACACACACCTTTTCAGAGCGGATGGCACCAGATCACCTATGTGGATTACTAC 1023

781 AGGAAGCAATCAACCAAGAGATCACCGACTTGAAGCAGCTCTTCTTGGTCAGCCAGCCC 840

1024 AAGCAGCAGTATGATATTACTGTATCGACCTGAATCAGCCCATGCTTGTGTAGTCTGTTA 1083

841 AAGAAAGCGCGGCGCTGGGGGAGACCTGCCAGGCGCTGCCATCTCATTTCTCAGCTC 900

1084 AAG---AAGAAAGAGAAATGACAAAGTGAAGGCTCAGCTCGGCCCACTGATACCTGAGCTC 1140

901 TGCTATCTTACAGGTTCACTAATGATAAATAATGCGCTAATGATTTTAAAGTATGAAGACTTA 960

1141 TGCTTTCTAAAGGGCTGACTTGACCCAGCAACATCTGATTTCCAGCTGATGAAAGCTGTG 1200

QY	2041	GTAGAGTTTACACGACCAAGATGGTATGACTTTTTTATCGTGAGCCAGCGCTGTGAAAGT	2110
Db	2275	TCAGAAAGCAACACGTAAACGAATGGTATGACTTTTATCTGATCAGCCAGCTGGCGCTCCCG	2334
QY	2101	GGTAGTGTTCCTCCACACATTAACATGTCATCTATGCAACAGCGGCTGAAGCCAGAC	2160
Db	2335	GGAACTGTATGTCCTACCTACTATAAAGTCACTATGATGCAACGGCTTGAAGCCCGAC	2394
QY	2161	CACATACAGCGCTTGACCTTACAAAGCTGTGCCACATCTATTCAACTGGCCAGGTGTCA	2220
Db	2395	CATATGCAGAGACTTACATTCAAATTGTGCCACTGTACTCAACTGGCGGGCAGTAGTC	2454
QY	2221	CGTGTTCCTGCTCTTGCAGTAGTACGCCCAAGAGCTGGCTTTCTTTCTTGTGCCAGAGTAT	2280
Db	2455	AGTGTCCCGACGACCATGTGAGTATGCTCAAGAGCTGACCTTTCTGTGGCACAAGCATT	2514
QY	2281	CACAGAGAGCAATCTGTCACTGTCAAACCGCTTTTACTACCTCTAA	2328
Db	2515	CATAAGAACCCAGTCTGGAATTAGCCAACCATCTCTTCTACTCTGTGA	2562
RESULT 10			
AK093475			
LOCUS			
DEFINITION	AK093475	3138 bp mRNA linear PRI 15-JUL-2002	
		Homo sapiens cDNA FLJ36156 fis clone TESTI202546, weakly similar	
		to Mus musculus mRNA for MIWI (piwi).	
ACCESSION	AK093475		
VERSION	AK093475.1	GI:21752360	
KEYWORDS		oligo capping; fis (full insert sequence).	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
REFERENCE		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS		1	
		Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S.,	
		Ono,Y., Hotta,T., Hiraoka,S., Murakawa,K., Takiguchi,S.,	
		Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,M.,	
		Yamashita,H., Chiba,Y., Sugiyama,T., Irie,R., Otsuki,T., Sato,H.,	
		Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y.,	
		Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y.,	
		Sekine,M., Kikuchi,H., Kanda,K., Magatsuma,M., Takahashi-Fujii,A.,	
		Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,	
		Nagahari,K., Masuno,Y., Nagai,K. and Isogai,T.	
		NEDO human cDNA sequencing project	
TITLE		Unpublished	
JOURNAL		2 (bases 1 to 3138)	
REFERENCE		Isogai,T. and Yamamoto,J.	
AUTHORS		Direct Submission	
TITLE		Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7	
JOURNAL		Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan	
		(E-mail:genomics@hri.co.jp, Tel.81-438-52-3975, Fax:81-438-52-3986)	
COMMENT		NEDO human cDNA sequencing project supported by Ministry of	
		Economy, Trade and Industry of Japan; cDNA full insert sequencing:	
		Research Association for Biotechnology (RAB); cDNA library	
		construction: Helix Research Institute (HRI) (supported by Japan	
		Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,	
		HRI, and Biotechnology Center, National Institute of Technology and	
		Evaluation; clone selection for full insert sequencing: HRI and	
		RAB; annotation: HRI and RAB.	
FEATURES		Location/Qualifiers	
source		1..3138	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="TESTI202546"	
		/tissue_type="testis"	
		/clone_lib="TESTI2"	
		/note="cloning vector: pME18SFL3"	
CDS		152..2710	
		/note="unnamed protein product"	
		/codon_start=1	
		/protein_id="BAC04179.1"	

/translation="MSGRARVARGIARSPSATEVGRIOASPLPRSVLNNNEASSN
 GFLGSRSTNDKYGISSDAGSTFMERGVKNDQFMDLSICTREKLAHVRNCKTSS
 GIPVLKVLNLFNLPDQWOLQYHVTYIPDLASRLRIALLYSHSELNEAKADGA
 ILFLSKLEKVELTSESTETIKMTITLAKELPSSPVCIOVFNIIFRKILKLS
 MYQIGRNFYNPSEPMKIPKHLSPGFAISVYPERKLLPSADVSXVLRNETVLEF
 MIALQRTGLSFTQCEKOLIGLIVLIRYNNRTYSIDDDWSVAPTHFQKRGTEI
 TVIVYQOYDITVSDNLPMPLVLLKKRNDNSEQAHLLEPELFCFTGLTDQATSD
 FQMKVAEKTLSPSRQORLARVDNIQRNTNARFELETWGLHFGSQISLITGRIVP
 SEKILMQDHICOPVSAADSKDIRCTKIILNAQSLNLTWILFICSDRTEYVAESFLNCLR
 VAGSGFVNDYPKIILKQBNPAFAIRAOYVDVQVLMVCLIPSNOKTYDYSIKKYL
 SSDCPVPSOCVLIARTIKVQNMMSLATKIAMOMTCKLAGELWAVSILPLKSLMVGVGDV
 CXDALSXDVMVGVCVASVNPRIIRFESRILQRTWIDVADCLKVFWTGLANKWYKXNH
 DEPARLIVTRAGVGQDLTLEIYEVPLLSVAGSGNSSTSRLSVIVVRKCMRPRFF
 TENNRTVONPLIGTWNBNWDFYLLISQVACRGTVSPTYNVIYDNDGLKPDH
 MORLTFKLCHLYNFWGIVSPAPQYAHKLTLFVAQSIHKPSELELANHFLY"
 BASE COUNT 912 a 678 c 714 g 834 t
 ORIGIN

Query Match 29.2%; Score 680; DB 9; Length 3138;
 Best Local Similarity 56.7%; Pred. No. 3.7e-172;
 Matches 1319; Conservative 0; Mismatches 1000; Indels 9; Gaps 3;

QY	1	ATGATCTTTGGGTGAACAACAGGAGAACTTAGACCATGTTAAAGAAATCAAAAACAGGT	60
DB	392	ATGGATTTGAGTATCTGTACAGAGAAAATTGGCACATGTGAGAAATTGTAAACAGGT	451
QY	61	TCTTCAGGCATTATAGTAGGTTAAGCACTAAACATTTCCGGGTGACATCCCGTCCCCAG	120
DB	452	TCCAGTGAATACCTGTGAACTGGTTACAAACCTCTTTAACTTAGATTTCCCCAAGAC	511
QY	121	TGGGCTTTATACGATACATTTGACTATACCCACTGATGAGAGCCAGAGACTCCGT	180
DB	512	TGGAGCTATACAGTACCATGTGACATATATTCAGATTTAGCACTAGAGGCTGAGA	571
QY	181	TCAGCTCTCTTTTCAACAGAGACTTAATTTGGAAAGTGTCATGCTTTTGTGGAACG	240
DB	572	ATTGCTTTTACTTTATAGTCACTAGTAACTTTCCACGAGGAAAGCAATTCAGCGTGCC	631
QY	241	ATATATTATTTACCTAAAAGACTACAGCAAAAGGTTACTGAAAGTTTATAGTAAGCCGG	300
DB	632	ATCTTTTCTGTCACAAAGCTAGAAGAAAGGTCACAGAGTTGTCAAGTGAACCTCAA	691
QY	301	AATGAGAGAGATGTGAGGATACGATCACTTTAAACAAATGAACCTTCCACTACACCA	360
DB	692	AGAGGTGAGACTATAAGATGACTATCACCTGAAGAGGAGCTGCGATCAAGATTCTCC	751
QY	361	ACTTGTTTCAGTCTTATATATTTTTCAGGAGGCTTTTGAAGATCATGATTTGCAAA	420
DB	752	GTGTGATCCAGGCTTCAATATCATCTTCAGAAAGATCTCAAAGATTGTCCATGTAC	811
QY	421	CAAAATTGAGCAAAATTTATTAACCCAAATGACCCAAATTTGATTTCCAAAGTACAGGTTG	480
DB	812	CAAAATTGAGCGAACTTCTATAATCTTCAGAGCCAAATGGAATTTCCCGCAGCAAAATTA	871
QY	481	GTGATTTGCGCTGGCTTCTACTTCCATCTTCAGTATGAAAACAGCATCATGCTCTGC	540
DB	872	TCCCTTTGGCGCTGGGTTTGGCATTTCTGTGCATATTTTGAAGGAGCTCTCTGTTAGT	931
QY	541	ACTGAGCTTAGCCATAAAGTCTCTCGAAGTGAGCTGTTTGGATTTCAATGTTCAACTTT	600
DB	932	GCTGATGTGAGTTACAAAGTCTCCGGAATGAGACGGTTCTGGAATTCATGCTGCTCTC	991
QY	601	TATCATCAGACAGAAACATAAATTTCAAGAACAAAGTTTCCAAAGAACTAATAGGTTTA	660
DB	992	TGTCAAAGAACTGGCTTGTCTCTTTTCAACCCAGCGTGTGAGAAAGCAGCTAATAGGGCTC	1051
QY	661	GTGTGTTCTTACCAAGATTAACATATAGACATACAGAGTGGATGATATTCAGCTGGGACAG	720
DB	1052	ATTGCTTTTACAGATACAAATACAGAACTTACTCCATTTGATGACATTCAGTGGTCAGTG	1111
QY	721	AATCCCAAGAGCACCTTTTAAGAAAGCCGAGCGGTCTGAAAGTCAAGCTTCTTAGAATACTAC	780
DB	1112	AGCCCAACACACACCTTTTCAGAAAGCGGGATGGCACCGGATCACCTATGTGATTACTAC	1171

QY	781	AGGAAGCAATCAACCAAGAGATCACCGACTTGAAGACGCTGTCTTGGTCAGCCAGCCC	840
DB	1172	AAGCAGCAGTATGATPATTACTGTATCGGACCTGAATCAGCCCATCTGTTAGTCTGTTA	1231
QY	841	AAGAGAAGCGGGGCCCTTGGGGGAGACCTGCCAGGCGCTGCCATGCTCATCTCAGCTC	900
DB	1232	AAG---AAGAAGAGAAATAGCAAACAGTGAGGCTCAGCTCGGCCACTGATACCTGAGCTC	1288
QY	901	TGCTATCTTACAGGTTCAACTGATAAAATCGTAAATGATTTTAACTGATGAAGACTTA	960
DB	1289	TGCTTTTAAAGGGCTGACTGACCGGCAACATCTGATTTCCAGCTGATGAAGCTGTG	1348
QY	961	CGCGTTTCATCAAGACTAACTCCAGAGCAAAAGGCGAGCGTGAAGTGGGACGACTCAATGAT	1020
DB	1349	GCTGAAAAGACACGCTCTCAGTCCCTTTCAGGCGGAGCAGCGCTTGGCCAGGCTTGTGAC	1408
QY	1021	TACATTCATAAARACGATTAATGTTTCAAAGGAGCTTCAGAGCTGGGTTTGAAGCTTTGAT	1080
DB	1409	AACATCCAGAGGAATACCAATGCTCGCTTGAACCTAGAGACTTGGGGACTGCAATTTTGA	1468
QY	1081	TCCAACTTACTGCTCTCTCAGGAAGAATTTTGCAAAACAGAAAGATTCACCAAGGTGA	1140
DB	1469	AGCCAGATA---TCTCTGACTGGCGGATTTGCGCTTCAGAAAANAATTAATGCAAGAC	1525
QY	1141	AAAAATTTGATTACAATCCCAATTTTCAGATTTGGTCCAAAGAAACAGAGGTGCACCA	1200
DB	1526	CACATAT--GTCAACCTGTGTCTGCTGCTGACTGCTCCAAAGGATATTTCGAACTTGAAG	1582
QY	1201	TTAATAGTGTTAAGCCACTAGATAACTTGGCTGTGTGATCTATACCGGAAGAAATTAATGA	1260
DB	1583	ATTTTAAATGCACAGTCTTTGAATACCTGGTGTGATTTTATGTAGCGACAGAACTGAATAT	1642
QY	1261	GCAGCAATTCATTGATACAAATCTATTTAAAGTTTACACGAGCATGGGATGCAAAATG	1320
DB	1643	GTTCGCGAGAGCTTCTGAACTGCTTGAGAGAGTTTGAGGTTCCATGGGATTTAATGTG	1702
QY	1321	AGAAAGCAATTAATGATTGAAGTGAAGTACAGAACTGAAGCCTACTTAAGAGTCTTTACAG	1380
DB	1703	GACTACCCCAAAATCATAAAGTACAGAAATCCAGCTGCATTTGTTAGAGCTATACAG	1762
QY	1381	CAAAAGTTCACAGCAGACACCCAGATAGTTGTCTGTCTGTGTTGTAAGTAATCGGAGAC	1440
DB	1763	CAATATGTTGATCTCTGATGTTTTCAGCTGTTAATGTGCATTTCTGCTTCTTAATCAGAAGACC	1822
QY	1441	AAATACGATGCTATTAAAAAATACCTGTGTACAGATTCCTTACCCCAAGTCAGTGTGTG	1500
DB	1823	TATTATGATTCATTAATAAATATTTGAGCTCAGACTGCCAGTCCCAAGCCAAATGTGTG	1882
QY	1501	GTGGCCGAACTTTAGGCAACAGCAAACTGTCAATGGCCATTTGTCTACAAAGATTGCCCTTA	1560
DB	1883	CTTGTCGGACCTTGAATAAAGCAGGGCATGATGATGATGATGATGATGATGATGATGATG	1942
QY	1561	CAGATGAATCCAGATGGGAGGAGCTCTGGAGGGTGGACATCCCTCGAGCTGCTG	1620
DB	1943	CAGATGAATCCAGCTGGAGGGAGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTGGGCTGTG	2002
QY	1621	ATGATCGTTGGCATCGATTGTTTACATGACATGACAGCTGGGCGGAGGTCAATTCACAGGA	1680
DB	2003	ATGTTGGTTCGTTATGATGTTCTGTAAAGATGCATCTCAGCAAGGAGCTGATGTTGTTGGA	2062
QY	1681	TTTGTTCAGCATCAATGAAGGGATGACCCGCTGTTTCTCAGCTGCTGATATTTTCAGGAT	1740
DB	2063	TGCGTGGCCAGTGTAAACCCAGAAATCACCAGGTGTTTCCCGCTGTATCTCTCAGAGA	2122
QY	1741	AGAGGACAGGAGCTGTGATGGGCTCAAGCTCTGCTGCAAGCGGCTCTGAGGCTTGG	1800
DB	2123	ACAACTGATGTTGCGAGATTGCTTGAAGTTTTCATGACTGAGAGCTCAACAAATGG	2182
QY	1801	AATAGCTCAATGATACATGCCAGCGGATCATCGTGTACCGGATCGGCTTAGGAGAC	1860
DB	2183	TACAGTCAATCATGATTTGCCAGCAGCGAATAATTTGTTACCGTCTGGTGTAGGGAT	2242

```
QY 1861 GCCAGCTGAAACACACTGGTGAACCTACGAAGTGCACAGCTTTTGGATTGTCTAAATCC 1920
Db 2243 GGTAGCTGGAACACACTTATTGAATATGAAGTCCACAGCTCTGAGCAGTGTGCAGAA 2302
QY 1921 ATTGTAGAGGTTACAACCTAGACTTAACCGTAATTTGGTGAAGAAAGAGTGAACACC 1980
Db 2303 TCCGGCTCAAAATACCAGCTCAAGACTGTGGTGAATTTGGTTCAGGAAGAGTGCATGCCA 2362
QY 1981 AGATTTTGTCTGCTGGAGGAGACTTCAGATCCACTCTCTGGAAACAGTTATTGAT 2040
Db 2363 CGATTTCTTACCGAAATGAACCGCACTGTACAGAACCCGCCACTTGGCACTCTTGTGGAT 2422
QY 2041 GTAGAGGTTACAGACCAAGATGGTATGACTTTTTTATCGTCAGCCAGGCTGTGAGAAGT 2100
Db 2423 TCAGAAGCAACACGTAACGAATGGTATGACTTTTATCTGATCAGCCAGTGTGCTGCCGG 2482
QY 2101 GGTAGTGTCTTCCACACATTAACATGTCATCTATGACAAACAGCGGCTGGAAGCCAGAC 2160
Db 2483 GGAACCTGTTAGTCTTACCTACTATATAATGTCATCTATGATGACAAACCGCTTGAAGCCGAC 2542
QY 2161 CACATACAGCGCTTGACCTACAGCTGTCACCACTTATTAACACTGTCGACAGTGTCAAT 2220
Db 2543 CATATGCAGAGACTTACATTCAAATTTGTCACCTGTACTACAACTGCGCGGCATAGTC 2602
QY 2221 CGTGTCTCTGCTCTTGGCCAGTACGCCCAAGAGCTGGCTTTTCTTGTGGCCAGAGTATT 2280
Db 2603 AGTGTCCCGACCACTGTGATGATGCTCAGAGCTGACCTTTCTGTGGCCAAAGCAATT 2662
QY 2281 CACAGAGACCAAACTGTGCTACTGTCAACCGCGCTTTACTACTCTAA 2328
Db 2663 CATAAAGAACCCAGTCTGGAATTAGCCACCACTCTCTTCTACTCTGTA 2710

RESULT 11
AK096837
LOCUS Homo sapiens cDNA FLJ39518 fis, clone PUAE2001188, moderately
DEFINITION similar to Mus musculus mRNA for MIWI (piwi).
ACCESSION AK096837
VERSION AK096837.1 GI:21756416
KEYWORDS oligo capping; full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M.,
Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
NEO human cDNA sequencing project
Unpublished
2 (bases 1 to 2456)
Isogai,T. and Yamamoto,J.
Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team) ; 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
Location/Qualifiers
1..2456
/organism="Homo sapiens"
```

```
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PUAE2001188"
/cell_type="pulmonary artery endothelial cells (HPAEC)"
/clone_lib="PUAE2"
/notes="Cloning vector: pME18SPL3-Primary culture,
pulmonary artery endothelial cells"
28..2028
/notes="unnamed protein product"
/codon_start=1
/protein_id="BAC04873.1"
/db_xref="GI:21756417"
/translation="MTITLKEPSSPVCIQVENILPRKILKLSMYQIGNFVNPFS
EPMEIPQKLSLWFGFAISVSYPFKLLFSADVSKYLNRNTELEFMTALCORTGLSC
TVCCEKGLIGLIVLIRNNRTYSIDIDMSVKPTHTPLKRDGTEITTVYIKQYDI
TSDLNQPMVLVSLKKRDNSEPOLAHLIPELCLTGLTDOATSDFOGMKAVAKTR
PVSQRQRIARLVNDINRNARFELTWGLHFGSGISLTCRIVPSKILMQDHICQ
LPSAASDKOIRTKILNALSNTWLILCSDRTEYVAESFLNCRVRSVSGMGNVDYP
KIIVQENPAAFVRAIOQYVDPQVLMCILPNSKTYVDSIKVLSDDCPVBSOVL
ARTLNKGMMMSIATKIAMQNTKLGELWAVEIPLSKLVVVGIDVCKDALSKDVNV
GCVASVPRITRWFSCILQRTMDVADCLKVFMTGALNKWKYNHNDLPARIIVTRAG
VGDQLKTLIEYEPQLSSVAESSNNTSSRLSVIVVRKCKMPFPFTEMMNTPNPL
GTVVDSEATRNWDYFLLISQVACRGTSPTTYNVIYDNGKLPKDPHMLRQLTFLKCHLY
YNWPGIVSPAPCOYAHKLTFLVAQSITKKEPSLELANHLFYL"
BASE COUNT 718 a 519 c 550 g 669 t
Query Match 26.5%; Score 616.8; DB 9; Length 2456;
Best Local Similarity 57.5%; Pred. No. 4.4e-155;
Matches 1170; Conservative 0; Mismatches 857; Indels 9; Gaps 3;
QY 293 AGACCGGAATGGAGAGGATGTGAGGATAAGCATCACTTTAACAATGAACCTTCCACCTA 352
Db 2 AACTCAAGAGGTGAGACTATAAAGATGATATCACCCTGAAGAGGAGCTGCCATCAA 61
QY 353 CATCACCAACTTGTGTGAGTCTTATAATATATTTTTCAGAGGCTTTTGAATAATCATGA 412
Db 62 GTTCTCCGTGTGCATCCAGTCTTCAATATCATCTTCAGAAAGATCTCAAAAAGTTGT 121
QY 413 ATTTGCAACAATTTGGAGAAATTTATTAACCAAAATGACCAATTTGATTTTCAAGTC 472
Db 122 CCATGTACCAAAATTTGACGGAATCTTATAATCTTTCAGAGCAATGGAATTTCCCGAC 181
QY 473 ACAGTTGGTCAATTTGGCTGGCTTCACTACTTCCATCTTCAGTATGAAACACGATCA 532
Db 182 ACAAATTATCCCTTGGCTGGCTTTCGCTTCTGTCATATTTTGAAGGAGAGCTCC 241
QY 533 TGCTCTGCACCTGAGCTTAGCCATAAAGTCTTTCGAAGTGAGAGTGTTTTGGATTTTCATGT 592
Db 242 TGTATTAGTGTGATGTGAGTTTCAAAAGTCTTCCGGAATGAGACGGTTCTGGATTTATGA 301
QY 593 TCAACTTTTATCATCAGACAGAGAAACATAAATTTCAAGAACAAAGTTTCCAAAAGAACTAA 652
Db 302 CTGCTCTCTGTCAAGAAAGCTGGCTTGTCTCTTTCACCCAGACGCTGTGAGAAGCAGCTAA 361
QY 653 TAGGTTTAGTTTGTCTTACCAAGTATACATATAGACATACAGAGTGCATGATTTGACT 712
Db 362 TAGGCTCATTGTCTCTTACAAGATACAAATACAGAACCTTCTCCATTTGATGACATTGACT 421
QY 713 GGGACCAAGATPCCCAAGACGACCTTTTAAAGAACCGCAGCGTCTTGAAGTTCAGTTCCTTAG 772
Db 422 GGTCAAGTGAAGCCACACACACACTTTCTTAAGCGGATGGACCCGAGATCACTTATGTGG 481
QY 773 AATCTACAGAAAGCAATACAAACAGAGATCAACCGACTTGAAGCAGCGCTCTCTTGGTCA 832
Db 482 ATTACTACAGCAGCAGTATGATATTACTGTATCGGACCTGAATCAGCCCATGCTTGTGA 541
QY 833 GCCAGCCCAAGAGAGCGGGCGCCCTGGGGGGAACACTGCGCAGGCGCTGCCATGCTCATTC 892
Db 542 GTCTGTGTTAAAG---AAGAAGAGAGAAATGACACAGTGCAGCTCAGCTCGCCACCTGATAC 598
QY 893 CTGAGCTCTGCTATCTTACAGGCTCTAATCTAATAAATGCGTAATGATTTTAACTGATGA 952
```



```

Dd 599 CTGAGCTCTGCTTTCTAAGAGGCTGACTGACAGGCAACATCTGATTTCCAGCTGATGA 658
Qy 953 AGAGCTTACCGTTTCATCAAGACTAACTCCAGAGCAAAAGGACGCGTGAAGTGGACGAC 1012
Dd 659 AGGCTGTGCTGAAAAAGACACGCTCTCAGTCCCTCAGGCGGACGACGCGCTGGCCAGGC 718
Qy 1013 TCATTGATTACATTCATATAAAGGATAATGTTCAAGGGAGCTTCGAGACTGGGTTTGA 1072
Dd 719 TTGTGGACAACATCCAGAGGAATACCAATGCTGCTGCTTGAACCTAGAGACTGGGACTGC 778
Qy 1073 GCTTTGATTCACACTTACTGCTCTCTCAGGAGAAATTTTGCAAAACAGAAAAGATTCAAC 1132
Dd 779 ATTTTGGAGCCAGATA---TCTCTGACTGGCGGATTTGCTTTCGAAAAAATATTA 835
Qy 1133 AAGGTGGAAAAACATTTGATTAACAATCCAAATTTGACAGATTTGTCCTCAAGAAAACAGAG 1192
Dd 836 TGCAAGACCACATAT---GTCACACTGCTGCTGCTGACTGGTCCAGGATATTCGAA 892
Qy 1193 GTGCACCATTAATTAGTGTAAAGCACTAGATAAATGGGCTGTGATCTATAGCGAAGAA 1252
Dd 893 CTGCAAGATTTTAAATGACACAGTCTTTGAATACCTGCTGCTGATTTTATGTAGCGACAGAA 952
Qy 1253 ATTATGAGCAGCAATTCATTGATACAAAACTATTTAAAGTTTACACGAGCATGGGCA 1312
Dd 953 CTGAATATGTTGCCGAGAGCTTTCTGAATGCTTGGAAAGATTACAGTTTCATGGGAT 1012
Qy 1313 TGCAATAGAAAAAGCAATTAATGATTGAAGTGGATGACAGAACTGAAGCTACTTAAAGAG 1372
Dd 1013 TTAATGTGACTACCCCAAAATCATAAAAAGTACAGAAAAATCCAGTGCATTTGTTAGAG 1072
Qy 1373 TCTTACAGAAAAAGTCAACAGACAGACACCCAGATAGTGTCTGTCTGTGTCAGTAATC 1432
Dd 1073 CTATACAGCAATATGTTGATGCTGATGTTTCACTTGGTAAATGTCATCTGCCCTTCTAATC 1132
Qy 1433 GGAAGACAAATACAGATGCTATTAATAAATACTGTGTACAGATGCCCTACCCCAAGTC 1492
Dd 1133 AGAAGACCTATTATGATTCATTAATAAATAATTTGAGCTCAGACTGCCAGTCCCAAGCC 1192
Qy 1493 AGTGTGTGTGCGCGCAACTTAGGCAAAACAGCAAACTGCTATGCGCAATTTGTCACAAAGA 1552
Dd 1193 AATGTGTGTGCTCGGACCTTGAATAAACAAGGCGATGATGATGATATCGCCACCAAGA 1252
Qy 1553 TTGCCCTACAGATGACTCCAGATCGGAGGAGAGCTCTGGAGGTGAGACTCCCGCTGA 1612
Dd 1253 TCGCTATGAGATGACTTCCAGAGCTCGGAGGAGAGCTGTGGGCTGTGGAATATACCTTTAA 1312
Qy 1613 AGCTCGTATGATCGTTGGATCGATTTGTATCATGACATGACAGCTGGGCGGAGTCAA 1672
Dd 1313 AGTCCTGATGTTGTCGTATGATGCTGTAAAGATGCACTCAGCAAGGACGTGATGG 1372
Qy 1673 TCGCAGGAATTTGTCGACGATCAATGAAGGAGTACACCGCTGGTTCTCAOGCTGCATAT 1732
Dd 1373 TTGTTGGATGCGTGGCCAGTGTAAACCCAGAAATCACCAGGTGGTTTCCCGCTGTATCC 1432
Qy 1733 TTCAGGATAGGACAGGAGCTGTAGTGGCTCAAGCTGCTCCTCGCAAGCGCTCTCA 1792
Dd 1433 TTCAGAAACAATGACTGATGTTGAGATGCTTGAAGATTTTCATGACTGAGACATCA 1492
Qy 1793 GGGCTTGGAAATAGCTGCAATGATGATCATCCAGCGGATCATCGTGTACCCCGATGGCG 1852
Dd 1493 ACAATGGTACAGTACATCATGATTTGCCAGCAGGATATTTGTACCGTGTGTTG 1552
Qy 1853 TAGGAGCGCGAGCTGAAAAACACTGGTGAATACAGAACTGCAAGTGGCAAGTTTTTGGATTGTC 1912
Dd 1553 TAGGGGATGGTCAAGTGAACACTTATTAATGATTAAGTCCACAGCTGCTGAGCAGTG 1612
Qy 1913 TAAATCCATCTGTAGAGTTACAACTTAGCTAACGTAACGTAATTTGGTGAAGAAAGAG 1972
Dd 1613 TGGCAGAAATCCAGCTCAAAATACAGCTCAAGACTGTCGGTGAATTTGGTCAGAAAGAGT 1672
Qy 1973 TGAACACCAAGATTTTTTGTCTGAGTCTGGAGGAGAGCTTCAGAAATCCACTTCTCGAAACAG 2032
Dd 1673 SCATGCCAGATTTCTTACCGAAATGAACCGCACTGTACAGAAACCCCACTTGGCACTG 1732
```

```

Qy 2033 TTAATGATGTAGAGTTTACCAGACAGAAATGATAGCTTTTATCTGTCGACCGAGGCTG 2092
Dd 1733 TTGTGATTCAGAAAGCAACAGTAACGAATGATGATGACTTTTATCTGATCAGCAGGTGG 1792
Qy 2093 TGAGAGTGTAGTGTCTCTCCACACATTAACAATGATCTATGACAAACAGCGCCTGA 2152
Dd 1793 CTTCCGCGGGAACCTGTAGTCTCTACTACTATAATGTCATCTATGATGACAAACGCTTGA 1852
Qy 2153 AGCCAGACCAATACAGCGCTTGACCTCAAGCTGTGACCATCTATTACAACTGGCCAG 2212
Dd 1853 AGCCGACCAATATGAGAGACTTACATTAATTTGTCACACTGACTACTAACTGGCCGG 1912
Qy 2213 GTGTCAATTCGTTCTCTCTCTCCAGTACGCCCAAGCTGGCTTTTCTTGTGTGCC 2272
Dd 1913 GCATAGTCAGTGTCCAGCACCATGTGATGCTCACAAGCTGACCTTTCTGTGGTGCAC 1972
Qy 2273 AGAGTATTCAGAGAGCGCAATCTGTCTACCTGTCAAACGGCTTTTACTACCTCTAA 2328
Dd 1973 AAAGCATTCATAAAGAACCCAGTCTGGAATTAGCCACCATCTCTTCTACCTGTGA 2028

RESULT 12
AY014899          3057 bp      mRNA      linear      INV 02-JAN-2001
LOCUS      Strongylocentrotus purpuratus seawi mRNA, complete cds.
DEFINITION      AY014899
ACCESSION      AY014899
VERSION        AY014899.1  GI:12007639
KEYWORDS
SOURCE
ORGANISM      Strongylocentrotus purpuratus
                Strongylocentrotus purpuratus
                Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
                Echinoidae; Euechinoidae; Echinacea; Echinoida;
                Strongylocentrotidae; Strongylocentrotus.
REFERENCE
AUTHORS      1 (bases 1 to 3057);
                Rodriguez,A.J. and Bonder,E.M.
TITLE        Seawi - Cloning and Molecular Characterization of a Sea Urchin
                Homolog of Piwi
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 3057)
AUTHORS      Rodriguez,A.J. and Bonder,E.M.
TITLE        Direct Submission
JOURNAL      Submitted (28-NOV-2000) Biological Sciences, Rutgers
                University-Newark, 101 Warren Street Room 135, Newark, NJ 07102,
                USA
FEATURES
            Location/Qualifiers
                1..3057
                    /organism="Strongylocentrotus purpuratus"
                    /mol_type="mRNA"
                    /db_xref="taxon:7668"
                    181..2745
                        /codon_start=1
                        /product="seawi"
                        /protein_id="AAG42533.1"
                        /db_xref="GI:12007640"
                        /translation="MDRRPGEQAPFPMQQAAPPQOASGMPVPGRSRGSTTQPGP
                        AVPEQREGLTSQIAKLAGSEGVGAVQDRRRRDEIELEPTRPENFKQAI
                        AGDIALIANGFKLTPQWLYQYRVDPEPEILNPRFALLKSHALLKGLTDM
                        DTYSLFLAEKVTLSAERKDSNVNICVTHVATLNPMTLHLNVLPRCLKMI
                        NMEQVGNVDPATAIDIKOHLQLPWFVTSILOEYDVMLLSDISHKVLTQTVWE
                        VMDFPNKAGRFKEETKLMIGQIVLTNNKTYRVDDIDFTTPADTFETRGPVS
                        YVDYFKSIRVITHVNQPMIVSRPKREKKGVGPAFLPELCTFLTDDMRADF
                        VMKMGQRTVRGPDRCRTLSGFTKLSNEEVKTYLDSWGFEDKQVLTGRLPP
                        EKLFPQKQSYNPSNADRSRTGNALTDALNNWKI FYTRDANRGQDFIKSLVR
                        VANPMGNVRGPEIVELPDDRTETVTRSLQAIADDTQIVVVIILPTNRKRDYDAI KKT
                        CVVTHPCSOVIVSRITLSKQMLNSVATKIANQNCMGGLWVEIPLSNMLIGID
                        SYHSLTKRSVLGTFVSMNKSQTSFFSCAFQAQGFAGNLSTLNNALNKRYIQIN
                        EKFFERLIIFRDVGSDQVNLVDYELKQIDTLDKVPQGTVEKLAIVVVDYTKLKP
                        FFANRLGGLNPPGTVIDDVTKPLHYDIFIISQSVRQGSVTSNVVVDYTKLKP
                        DHMQLTKLTHLYFNWPGTVRVPAPQWAKLAFLIQSVHEVNNRLADTLFV"
                822 a      806 c      796 g      633 t
BASE COUNT
ORIGIN
```

Query Match 24.4%; Score 568.6; DB 3; Length 3057; Best Local Similarity 54.9%; Pred. No. 4.8e-142; Matches 1233; Conservative 0; Mismatches 994; Indels 18; Gaps 5;				
Qy	89	CTAACCAATTCGGGTGATCCCGTCCCGAGTGGCCCTTATATACAGTATCACATTTGACT	148	
Db	512	CAAAACGGCTTCAAGTTGAAGACTAAACAGACTGGCAACTCTACAGTACCGGTTGACT	571	
Qy	149	ATAACCACTGATGAAGCCAGACAGCTCCGTTACAGCTCTCTTTTCAACACAGAGATC	208	
Db	572	TTGAGCCCTGAGATCTCTGAACCTAGAGCTGGTTTGGCCCTCTGAAAGGGCCACAGTGTCT	631	
Qy	209	TAAITGGAAGTGTCTATGTTGATGAACATATTTTATCTTAAAGACTTACAGC	268	
Db	632	TTCTTTGTAAGGTTAAGCTCGATATGGACCACTATGTATCTTTTCAGCTAGCAG	691	
Qy	269	AAAAGGTTACTGAAGTTTATAGAACCCGGAATGGAGAGATGAGGATACAGATCA	328	
Db	692	AAAAGGTAACCAACCTGAGCGCAGAGAGGAAGGATGGGTCTAAACGTGAATATCTGTGTGA	751	
Qy	329	CTTTAAACAAATGAACCTCCACTCATACCAACTTGTTCAGTTCCTATATATTT	388	
Db	752	CTCATGTGTCTACTTTAAACCCCTAGGACCAACACACCCCTGCATCTCTACAGCTCTCT	811	
Qy	389	TCAGAGGCTTTTGAATAATCATGATTTGCAACAAATTTGACGAATTTATTAACCCAA	448	
Db	812	TTAGAAGGTGTTTGAAGATGATCAATATGGAGCAAGTTGGACCACTACTATGATCCTA	871	
Qy	449	ATGACCAATTTGATATTCAGATGACAGTTGGTGTATTTGGCTGGCTTCTACTTCTCA	508	
Db	872	CGGCCGCAATTTGACATCAAGCAGCATGGGCTGCAGCTGTGGCCAGGGTTCTGTGACATCCA	931	
Qy	509	TCCTTCAGTATGAACACAGCATGCTCTGCACTGACGTTAGCCATTAAGTCTCTCGAA	568	
Db	932	TCCTTCAGTACAGATGACAGTGTGCTGTCTGACATCAGCCACAGGTTCTGGCTA	991	
Qy	569	GTGAGACTGTTTGGATTTCAATTTTCACTTTTATCATCAGACAGAAACATAAATTC	628	
Db	992	CCGACCGTCTGGAGGTCATG---AATGATCTCTTCAACAAAGCTAGGGCAGTTCA	1048	
Qy	629	AAGAACAGTTTCCAAAGACATTAATAGGTTTAGTTGTTCTTACCAAGTATTAACATAGA	688	
Db	1049	AGAGGAATCACAAAGCTCATGATTTGACAGATTTGCTCTCACAAAGTACACACACAGA	1108	
Qy	689	CATACAGAGTGGATGATATGACTGGGACCAAGATCCCAAGAGCACCTTTAAGAAAGCCG	748	
Db	1109	CCTACCGGTGATGATATGATTTGATTTGATTTGTTCTTACCAAGTATTAACATAGA	1168	
Qy	749	ACGGCTCTGAAGTCAAGCTTTTGAATATCTACAGGAACATACACCAAGAGATCACCG	808	
Db	1169	CCGGACC---AGTTTCTATGTCGATTACTTCAAGAGTCTCTATGAGAGGTCATTACG	1225	
Qy	809	ACTTGAAGCAGCTGCTTGTGTGACGACCAAGAGAAAGGGGGCCCTGGGGGACAC	868	
Db	1226	ACGTGAACCAACCCATGCTGTGTCTCCGCCAAGAGAGAGGAGAAAGAGGG-----	1280	
Qy	869	TGCCAGGCTGCGCATGCTCATCTCTGAGCTCTGCTATCTTACAGTCTTAATGATAAAA	928	
Db	1281	-AGTTGTCGAGCTTACTACCGCTGAGCTGTGCTTCTGACTGGCCCTAAGCGACACA	1339	
Qy	929	TGCGTAATGATTTTAACTGATGAAGACTTTAGCCGTTTATACAGAGATAACTCCAGAGC	988	
Db	1340	TGCGCGCGATTTCAACGTCATGAAGGACATGGGCCAACACACTCTGTGTGACCCCAAG	1399	
Qy	989	AAAGCAGCGTGAAGTGGAGAGATCATNTGATTAATTCATTAATAAGATATGTTCAAA	1048	
Db	1400	ACAGATGTAGAACCGCTCAGCGGCTTCTATCAAAAGCTCAGCAGCAATGAAGAGTGAAGA	1459	
Qy	1049	GGGAGCTTCGAGACTGGGTTTGACTTTGATTTCAACTTACTTCTCTCAGGAAGAA	1108	
Db	1460	CTTACTAGATTCCTGGGGCATGGAGTTTCGACAAAGAGCAGGTCAAACTGCTGGCGGTG	1519	
Qy	1109	TTTTGCAACACAGAAAGATTACCAAGGTGGAAAAACATTTGATTACAAATCCACAATTTG	1168	

Db	1520	TCCTGCCACCTGAGAACTATTTCAGAGGGTAAACAGTTCAGTTACATCCGAGCATG	1579	
Qy	1169	CAGATTGGTCCAAAGAAACAAAGAGGTGCACTTAAATTTAGTGTAAAGCCACTAGATAACT	1228	
Db	1580	CGGACTGGTCAGAGACACAGAGGGTAAATGSCCTCACTGATGCAAGATCTCTCAACT	1639	
Qy	1229	GGCTGTTGATCTATACCGGAGAAATTTATGAAGCAGCCAAATTCATTGATACAAATCTAT	1288	
Db	1640	GGAAGATCTTCTACACACGCGCGGACGCCAACAGGGGCCAGGACTTCTAAGTCACTCG	1699	
Qy	1289	TTAAAGTTTACACAGCCATGGCATGCAAAATGAGAAAGCAATTAATGATTGAAGT---GG	1345	
Db	1700	TACGAGTCGCAACCCCAATGGTATGAACGTACAGGACACAGATAGTGAACCTTCCCG	1759	
Qy	1346	ATGACAGAACTGAAGCCTACTTTAAGATCTTTACAGCAAAAGGTTCACAGCAGACCCAGA	1405	
Db	1760	ATGATCGTACAGAGACCTACACAGATCCCTTACAGGCACAAATCGCACAGATACAGA	1819	
Qy	1406	TAGTTGCTGCTCTCTCAAGTAATCGAAGGACAAATACGATGCTTATTAATAATACC	1465	
Db	1820	TTGGTGGTGTGATCCTGCCAACGAAACAGAAAGGACCGGTAGCATGCCATCAGAGAGCT	1879	
Qy	1466	TGTGTACAGATTGGCTTACCCCAAGTCAGTGTGTGTGGCCGCAACCTTAGGCAAAACAGC	1525	
Db	1880	GCGTGTGACCCATCCATGCCAGCCAGGTCTCGTCTCCAGGACCTGTCTCAAGCAGC	1939	
Qy	1526	AACTGTTCATGGCCATTTGCTACAAAGATTGCTTACAGATGAACCTGCAAGATGGAGGAG	1585	
Db	1940	AGATGCTCATGTCTGTGTAGCCCAAGATGTCATGTCAGATGAACCTGCAAGATGGCGGAG	1999	
Qy	1586	AGCTTCGAGGGTGGACATCCCTCGAAGCTCGTGATGATGCTTGGGATCGATTGTTACC	1645	
Db	2000	ACCTTCGCGAGTGCAGATCCCGCTGACCAACCTGATGATCATCGTATTGACAGCTTACC	2059	
Qy	1646	ATGACATGACAGCTGGCGGAGGTCAAATCGCAGGAATTTGTTCCAGCATCAATGAAGGGA	1705	
Db	2060	ATGACTCTCTGACCAAGGCGCTCGCTCTGGGATTCGTGGCTTCCATGAACAAGTCTC	2119	
Qy	1706	TGACCCGCTGGTTCCTACGCTGCATATTTCAAGATAGAGACAGGAGCTGGTAGTGGCC	1765	
Db	2120	AAACAGCTTCTTCTCTAGTTGTGCTTCCAGCAGCAGCAGGAGAAATTCGTGCTAACCC	2179	
Qy	1766	TCAAAGTCTGCTCAAGCGCTCTGAGGGCTTGGAAATAGCTGCAATGAGTACATGCCCA	1825	
Db	2180	TGAGTACCTCATGAACAATGCTCCCTGAGCGATCTATCAGATCAACGAGAGTTTCTCG	2239	
Qy	1826	GCCGATCATCTGTATACCGGATGGCGTAGGAGACGCCAGCTGAAAACACTGGTGAAC	1885	
Db	2240	AGCGTATCATCTTCCGTGATGCTGTGTGACAGCCCAAGTCAATCTGCTGCTGACT	2299	
Qy	1886	ACGAGTGCACAGTCTTTTGGATTGTCTA---AATCCATTTGGTAGAGTTTACACCCCTA	1942	
Db	2300	ATGAGCTCAAAACAAATCAAGGACACCTTGGACAAAGTCTACCTCAAGGGACGGTTCACA	2359	
Qy	1943	GACTAACGGTAATTTGCTGAAGAAAAGAGTGAACACACAGATTTTTCCTCAGTCTGAG	2002	
Db	2360	AGCTGGCCGTTGTGCTGCTGAGAGCGATCAATACCGCTTCTTTGCCAATCTGAGAG	2419	
Qy	2003	GAGACTTCAGAAATCCACTTCTTGGAAACAGTTATTTGATTTAGAGGTTTACAGACCAAGAT	2062	
Db	2420	GAGSCCTGTCCAAACCTCCCTCCCTGGAACCTCATCGATGATGCTGCACCAAGCCCCACC	2479	
Qy	2063	GGTATGACTTTTATCGTGAGCCAGGCTGTGAGAGTGTGATGTTTCTCCCAACAATT	2122	
Db	2480	TCTATGACTACTTCTCATCATGTCAGAGTGTCCGCCAGGGCTCGTGTCCCCACGCTCT	2539	
Qy	2123	ACATGTCATATGACAAACAGCGGCTGAGCAGACCAATACAGGCTTGACCTTAC	2182	
Db	2540	ACACGTTGTATGACACCACTGGTCTCAAGCCAGACCAATGACAGAGTTGACCTACA	2599	
Qy	2183	AGCTGTGCCACATCTATTTACAACTGGCCAGGTGTCTTCTGTTTCTGCTCTCTGCGAGT	2242	

Db 2600 AGCTGACCATCTCTACTTCACTGCGGGGACAGGTGGTGTCTCTGCCCCGTGCATGT 2659
Qy 2243 AGCCCCACAAGCTGGCTTTTCTTGTGGCCAGAGATTTACAGAGAGCCAAATCTGTAC 2302
Db 2660 ATGCACACAAGCTGGCTTTTCTTGTGGCCAGAGGTGCATGAAGTGGCCCAACAGAC 2719
Qy 2303 TGTCAACCGCCCTTTACTACTCTTA 2327
Db 2720 TAGCCGACACCCCTCTCTATCTTA 2744

RESULT 13
AK116906
LOCUS AK116906 3458 bp mRNA linear INV 30-NOV-2002
DEFINITION Ciona intestinalis cDNA, clone:cieg052a12, full insert sequence.
ACCESSION AK116906
VERSION AK116906.1 GI:23589627
KEYWORDS FLI CDNA.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.

REFERENCE 1
AUTHORS Satou, Y., Yamada, L., Mochizuki, Y., Takatori, N., Kawashima, T.,
Sasaki, A., Hamaguchi, M., Awazu, S., Yagi, K., Sasakura, Y.,
Nakayama, A., Ishikawa, H., Inaba, K. and Satoh, N.
TITLE A cDNA resource from the basal chordate Ciona intestinalis
JOURNAL Genesis 33 (4), 153-154 (2002)
MEDLINE 22191024
PUBMED 12203911

REFERENCE 2 (bases 1 to 3458)
AUTHORS Satou, Y. and Satoh, N.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2002) Nori Satoh, Kyoto University, Department of
Zoology; Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
(E-mail: satoh@ascidian.zool.kyoto-u.ac.jp, Tel:81-75-753-4095,
Fax:81-75-705-1113)
COMMENT Ciona intestinalis cDNA Project (URL:
http://ghost.zool.kyoto-u.ac.jp/indexr1.html).
FEATURES
source
1..3458
/organism="Ciona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="cieg052a12"

BASE COUNT 1062 a 626 c 810 g 960 t

ORIGIN
Query Match 21.6%; Score 503.2; DB 3; Length 3458;
Best Local Similarity 54.3%; Pred. No. 2.3e-124;
Matches 1225; Conservative 0; Mismatches 998; Indels 33; Gaps 9;
Qy 89 CTAAACATTTCCGGCTGACATCCCGTCCCGAGTGGCCCTTATCAGTATCATTGACT 148
Db 625 CCAACTTTTCCATTTAGTTGCACAGGCGAGTGGTCTATACAGTATAGTTGACT 684
Qy 149 ATAAACCACTGATGAGCCAGAGACTCCGTTAGCTCTTCTTTTCAACAGAGATC 208
Db 685 ATGCACCTGAGTGCACACAGAGGCGAGGAAAGGAATGTAAGAATCATGAAGGCA 744
Qy 209 TAAATGGAAAGTGTCATGCTTTTGTAGTGAACATATATTTTACTTAAAGACTACAGC 268
Db 745 TCAITGGGAATGCTTTCATGTTTGTATGGAAACAATGTTTGTACCAACCAAGAGTGGCCG 804
Qy 269 AAAAGGTTACTGAAGTTTTTAGTAGACCG---GAATGGAGAGGATGTGAGGATAAGCA 325
Db 805 ATAAAGAAACAGTGGTTTACTCAAGCGTCAACGATGAGAGCAATGTACAGCTTACGA 864
Qy 326 TCATTTAAACAATGAATCTTCAACCTTACATCAACACTTTGTTGAGTTCTATATATTA 385
Db 865 TCGCATTAACAGAGCTTCCACCCAGTTCGCCATCTCAATAACAATATACAACATTC 924
Qy 386 TTTTCAGGAGGCTTTTGAAAAATCATGAATTTGCAACAATTTGGACGAAATTTATTAACC 445

Db 925 TATTGAGATGGTGTCTTTGATAAGATAGCATGAAGAGTGGAGGAATTTCTACAATC 984
Qy 446 CAATATGACCAATTTGATATCCAAAGTACAGG-----TTGGTGAATTTGGCTGCTTCA 499
Db 985 CTTCATGTCTATCTACATGCGAATGCCAGAGGTCAACTTTGAACATATGGCCAGATTCA 1044
Qy 500 CTACTTCATCTCTTCACTATGAAAAACAGCATCATGTCTCTGCACTGACGTTAGCCATAAG 559
Db 1045 TCACCAGTATTTCTGCAGTATGAGAAGAGTGTGTTGTTATGTCTGCTGAGTTTCCCATAGT 1104
Qy 560 TCCTTGAAGTGAGACTGTTTGTGATTTCACTTCACTTTTAT-----CATCAGACAG 613
Db 1105 TGATGCGGAAGATTCGGTGTCTGATATCTTACCGGAGATGATATGGGACTGTAGGCAAC 1164
Qy 614 AAGAACAATAATTTCAAGAACAACTTTCCAAAGAACTAAATAGGTTTGTGTTCTTACCA 673
Db 1165 GTGGAAGAGATTTTAAACAGAGATGGAATAATTTCTTGGTGGGACAAATTTCTGACAA 1224
Qy 674 AGTATAACAATAAGACATACAGATGAGATGATATGATGCGGACAGAAATCCCAGAGCA 733
Db 1225 GATACAAACAATAAGACATACCGCATAGATGGATTTGAATGGAACCTTAAATGTTAATGA 1284
Qy 734 CCTTTAAGAAAGCCGAGGCTCTGAAGCTCTGAACTCTTTAGATATCTACAGGAGCAATACA 793
Db 1285 AGTTTGAGCGAAATCTGGTTCC---GTCTGTGTATGGATTTATATAAGAGCAATACA 1341
Qy 794 ACCAAGAGATCACCGACTTGAAGCAGCGCTGTCTTGGTCAGCGAGCCCAAGAGAGGCGGG 853
Db 1342 ACATAGTATCCGAGATACAAACCAACCTCTGCTTCTATCTCTGCTTAAACAATCAGAGA 1401
Qy 854 GCCCTGGGGGACACTGCCAGGCGCTGCCAGTCTCATTTCTGAGCTCTGTCTATCTTACAG 913
Db 1402 TTCGAAAGGTGGATTTGGAGGTGTTTCA---CTTGTGCCGAGTTATGTACAGTGACAG 1458
Qy 914 GTCTAATCTGATAAATCGGTAATGATTTAACTGTGATGAAGAGCTTAGCCCTTCATACAA 973
Db 1459 GGCCTACTGTAGAGTGGCGGCTGACTTCAATACATGAAAAGGTGGCAGTCTACACAA 1518
Qy 974 GACTAACTCCAGAGCAAGGCGGTGAAGTGGGACGACTCATTTGATTACATTTATATAAA 1033
Db 1519 AACAAAGNCTTACCAAGAGAAACAAGCGCTGAAAAGCTTTATACAGCAATCACCACA 1578
Qy 1034 ACGATAATGTTCAAAGGAGCTTCGAGCTCGGGTTGAGCTTTGATTTCCAACTTACTGT 1093
Db 1579 AACTGAGTTGAAAAGAGATTTGCAGAGTGGGGCTTAGATTTGAGGATCGACTCTTTGG 1638
Qy 1094 CCTTCTCAGGAGAAATTTTGCAACAGAAAGATTCACCAAGGTGGAAAAACATTTGATT 1153
Db 1639 ATTTAAAGGGAAGAGTTCTTGACTCTGAAACCAATAATGTTTGGTGTATTAAGCAAGCAC 1698
Qy 1154 ACAATCCACATTTGCGAGATTTGGTCCAAAGAAAACAAGAGGTGCACCAATTAATTAGTTA 1213
Db 1699 AGAGTAGAGA---TGCACTTTGGGATCAGAGTTCCGAGACAGAGACTACTTAAGTGA 1755
Qy 1214 AGCCACTAGATACTGGCTGTGTATCTATACCGGAGAAATTTAGACGACCAATTCAT 1273
Db 1756 TTGATCTTTCAGAGTGGGCCATCTCTGTTTCTCTAGAGATAAAAGATGTGCGGAGGATT 1815
Qy 1274 TGATACAAATCTATTTAAAGTTACACAGCCATGGGCATCAAAATGAGAAAAACAATAA 1333
Db 1816 TTGAGAGAGTTGTTTAAAGTATCTGTAACTAGGATTCGGTGTGCTCGTCCGACTC 1875
Qy 1334 TGATTGAAGTGGATGACAGAACTGAAGCTTAAAGAGTCTTACAGCAAAAGTCT---A 1390
Db 1876 CTGTTGAGTTGGAATCGGATCAATGTTTCAGAGAGAGGATCCGGGATCTAATGGGAGGA 1935
Qy 1391 CAGCAGACACCCAGATAGTTCTGTCTGTGTCTGTCAAGTAACTCGGAAGCAAAATACGATG 1450
Db 1936 ATCGGAAGACACAGCTGATATGTTGTATGATGCCAGCAGTAGAAGAGACAGATATGAAG 1995
Qy 1451 CTATTAAAAATACCTGTGTACAGATTGCCCTCCCAAGTCACTGTGTGTGGTGGCCGAA 1510

Db	1396	GCATCAAGAAAGTTTGTGTGGATATGCGGTTTCCAAGTCAGGTTGTATTATCAAGCA	2055
Qy	1511	CTTAGGCAACACAAACTGTCATGGCCATGCTACAAAGATTGCCCTCAGATGAAT	1570
Db	2056	CATTATCAAAACCAACGAGTAATAGCATCGCTACAAAGATTGCAATCCAGATGAAT	2115
Qy	1571	GCAAGATGGGAGGAGAGCTCTGGAGGGTGGACATCCGCCCTCAAGCTCGTGATGATCGTTG	1630
Db	2116	GTAAGCTTGGTGGTGGAGCATGGCTGTCAAGATACCGTTAAAGGAATGTATGATCGTAG	2175
Qy	1631	GCATCGATTGTTACCATGACATGACAGCTGGCGGAGGTCAATCGCAGGATTTGTGCA	1690
Db	2176	GATTTCGATACATCCACGACTCATCACAGAAAGGAAGATCTGTTGGTGGTTTGTGGCA	2235
Qy	1691	GCATCAATGAAGGATGACCCGCTGTTCTCAGCGTGCATATTTTCAGGATAGAGCAGG	1750
Db	2236	GTATTAAATCAACTTACACAAAGATGGTACTCGAGTGTCACTTTCACATTTCAAGAGTTG	2295
Qy	1751	AGCTGGTAGATGGGCTCAAAAGTCTGCCCTGCAAGCGGCTCTGAGGCTTGGAAATAGCTGCA	1810
Db	2296	AATTAAATCGATTGATGAAACGCTGCAAGAGGGGCAATTCMAAAGTTTCATCAGGTGA	2355
Qy	1811	ATGAGTACATCCACGACGCGATATCGTGTACCGCGATGGCGTAGGACACGCGCAGCTGA	1870
Db	2356	ATGGAAACTACCCGCTCGTATTTGTCACTACAGAGATGGGTCGTTGATGGAACAGCTGG	2415
Qy	1871	AAACACTGGTGAATACGAAGTGCACAGTTTTCGATTGTCTAAA---ATCCATTGGTA	1927
Db	2416	CTGTTGTGAAGGACCATGAGATATCACAACTACTTGAATGTCTTAAGATACCAAGTATA	2475
Qy	1928	GAGGTTACAAACCTTAGACTAAGCGGTAATTTGGTGAAGAAAGAGTGAACACCAAGATTTT	1987
Db	2476	AAGGCTTCACTCCCAATGTTCTTTCATCGTGGTTAAGAGAGATTAAACGAGTTCT	2535
Qy	1988	TGCTCACTCGGAGGAA---GACTTCAGATTCGACTTCTTCTGCGAACACTTATGATGAG	2044
Db	2536	TCGCTGCTGGTCGACACCAAGCACTAATGAACGCACTCTGAGTACGATCATCGATGACA	2595
Qy	2045	AGGTTACAGACCAAGAGTATGATGATTTTATCGTGCAGCCAGGCTGTGAGAGTGTGA	2104
Db	2596	CGGTACAGAACCAAGATGGTATGATTTCTTTGTGTCAGCAGTCCGTACGAGAGGCA	2655
Qy	2105	GTGTTTCTCCACACATTAACATGTCATCTATGACACAGCGGCTGAGCCAGACACACA	2164
Db	2656	CAGTCAGTCCACGATTAACATGTTGTATATGACACATCACTCAACCTGATGACA	2715
Qy	2165	TACAGCGTGTGACCTACAGCTGTGCGCACATCTATTAACAATGCGCAGGTTGATTCGTG	2224
Db	2716	TGCAACGACTCACTTACAAAGTTGTGTACCTTTACTATAACTGGCCAGGAAACCATTCGTG	2775
Qy	2225	TTCTGCTCCTTGGCAGTACGCCACAGCTGGCTTTTCTTGTGGCCAGAGTATTCACA	2284
Db	2776	TACAGCCGCAATGCTATATGCGCACAGATAGCAATTTCTGTCGGAACAAAGTATTCATG	2835
Qy	2285	GAGAGCCAAATCTGTCACTGTCAAAACCGCTTTACT	2320
Db	2836	CACCTGTGTCTCTACTCTAGCCGACAACTTACT	2871

RESULT 14

AY014900

LOCUS

AY014900

DEFINITION

Strongylocentrotus purpuratus clone 6-2 seawi mRNA, partial cds.

ACCESSION

AY014900

VERSION

AY014900.1

GI:12007641

KEYWORDS

Strongylocentrotus purpuratus

SOURCE

Strongylocentrotus purpuratus

ORGANISM

Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Echinoidea; Echinoidae; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.

REFERENCE

1 (bases 1 to 2605)

AUTHORS

Rodriguez, A.J. and Bonder, E.M.

TITLE		Seawi - Cloning and Molecular Characterization of a Sea Urchin	
JOURNAL		Homolog of Piwi	
REFERENCE		Unpublished	
AUTHORS		2 (bases 1 to 2605)	
TITLE		Rodriguez,A.J. and Bonder,E.M.	
JOURNAL		Direct Submission	
		Submitted (28-Nov-2000) Biological Sciences, Rutgers	
		University-Newark, 101 Warren Street Room 135, Newark, NJ 07102,	
USA			
FEATURES		Location/Qualifiers	
.. source		1. .2605	
		/organism="Strongylocentrotus purpuratus"	
		/mol_type="mRNA"	
		/db_xref="taxon:7668"	
		/clone="6-2"	
		181. .>2605	
		/codon_start=1	
		/product="seawi"	
		/protein_id="AAG42534.1"	
		/db_xref="GI:12007642"	
		/translation="MDRRPGEQAPPNQQAAPQEAASSGMPVPGRRSRGGTTQPGP	
		AVPKQPRCLTSQIAKLAGSGGRVGAQDRRRRDEIELEPIITRENFKQAI	
		AGDKIALIANGFKLTKPDWLYQYRDFEILNPRARFALLKSHALLGKGLTLM	
		DMYSLFKAEKYNLSAERKGSNNVICVHTATLNPMAPTLHLYNLFRRCLKMI	
		NMEQVGRNYDPTAAIDIKOHLQLPFGFTVILQYDVMLLSDISLHKVLRITVWE	
		VMNDLENKARGREKEBITKLMIGOVLTKNNKTYRVDIDETTPADLPETRSQPV	
		YVDYFKXSIRVTHDYNQMLVSRPKRBEKGVGPAKLPELCLFLGLSDMPADEN	
		VMKMDGHTRVGFQDRCRTLSUGIFKLUSNEBKVLYLDSWGEFDEKQVLTGRLVPP	
		EKIUFQKQKSYNPSNADSRDRTGNALTDAKILANNKI FYTRDANRGQDFIKSLVR	
		VAMPFMNVRGPEIPELDDRTETYSLSQAQADTQIVVVLPTNKRDRDYDAIKKT	
		CVYTHPCPSOIVSRITLSKQMLMSVATKIAMQNCNMGDLNRVETPLNLMITIGID	
		SVHDSLTFKGRSVLGFVASMKNKTSPPSCAFQHAOCFAGNLSTMNALKRYQIN	
		EKXPERILIFRDGVDGSOVNLVNDYELKQIKDLDKIVYPOGTVHKLAVVVKRINR	
		FFANRLGSLNPPPGIVDDVVKPHLYDYFIISQSVRQSVSFTSNVYDTGLKE	
		DMORLYTGI"	
BASE COUNT		718 a	695 c 686 g 506 t
ORIGIN			
Query Match		21.5%; Score 499.4; DB 3; Length 2605;	
Best Local Similarity		54.1%; Pred. No. 2.4e-123;	
Matches 1135; Conservative		0; Mismatches 946; Indels 18; Gaps 5;	
Qy	89	CTAACCAATTTCCGGCTGACATCCCGTCCCGAGTGGGCTTATATCAGTATCACATTGACT	148
Db	512	CRAACGGCTTCAAGTTGAAGACTAAACACAGACTGGCAACTCTACCAAGTACCGGTTGACT	571
Qy	149	ATAACCCACTGATGGAGCCAGAGACTCCGTTACGCTCTTCTTTTTCACACAGAATC	208
Db	572	TTGAGCCTGAGATCCTGAACCCCTAGAGCTGTTTTCGCCCTCCTGAAGGGCCACAGTGCTC	631
Qy	209	TAATTGGAAAGTGCATGCTTTTGTAGGAACGATATATTTTACCTTAAAGACTACAGC	268
Db	632	TTCTTGTGTAAGGTTAAACGCTCGATATGGACACCATGTATCTCTTTTCAAGCTAGCAG	691
Qy	269	AAAAGGTTACTGAAGTTTTCATGTAAGACCCCGAATGGAGAGGATGTGAGGATAACGATCA	328
Db	692	AAAAGGTTAAACCACTGAGCGCAGAGAGGAGGATGGGTCTAAACGTGAATATCTGTGTGA	751
Qy	329	CTTTTAAACAATGAATCTTCCACCTACATCAACAACTGTTTTCAGTTCTCTATAATATTTT	388
Db	752	CTCATGTTCATCTCTTAAACCCCATGGCACCAACACCCCTGCACTCTCTACACGCTCTCT	811
Qy	389	TCAGGAGGCTTTTGAATAATCATGAATTTTCCAAACAAATTCGACGAAATTTATTAACCCAA	448
Db	812	TTAGAAGGTGTTTGAAGATGATCAATATGGAGCAAGTTGGACGCAACTACTATGATCTTA	871
Qy	449	ATGACCCCAATTGATATTTCCAAGTCAAGGTTGGTGTATTTGGCCTTGCTTCACTACTTCCA	508
Db	872	CGGCCGCCATTGACATCAAGCAGCATGGGCTGAGCTGTGGCCAGGGTTCGTGACATCCA	931
Qy	509	TCCTTCAGTATGAAAACAGCATCATGCTCTGCACTGACGTTAGCCATAAAGTCCTTCGAA	568
Db	932	TCCTTCAGTACGAGTACGAGCTGATGCTGTCTGCTGACATCAGCCACAGGTTCTCGGTA	991

QY	569	GTGAGACTGTTTGGATTTCATGTTCAACTTTTATCATCTCAGACAGAAGAACATAAAATTC	628
Db	992	CCGAGACCGTCTGGAGGTCATG---AATGATCTCTTCAACAAGCTAGGGCAGGTTCA	1048
QY	629	AAGAAACAAGTTTCCAAAGAACTAATAGTTTAGTGTGTTTACCAAGTATAACAATAAGA	688
Db	1049	AGAGGNAATCAANAAGCTCATGATTTGGACAGATGTCCTCANAAGTACACAACAAGA	1108
QY	689	CATACAGATGGATGATATTGACTGGACACAGAAATCCCAAGAGCACCTTTAAGAAAGCCG	748
Db	1109	CCTACCGGCTAGATGATATCGACTTTGACACTACTCCAGCGGTACTTTTGGAGACCGGTT	1168
QY	749	ACGGCTCTGAAGTCAGCTTCTTAGAATCTACAGGAAGCAATACAAAGAGATCAACG	808
Db	1169	CCGAGCC---AGTTTCCTATGTGCGATTACTTCAGAAAGTCTCTATGAGAGGTCATTCA	1225
QY	809	ACTTGAAGCAGCCGTCTTGCTGAGCCAGCCCAAGAGAAGCGGGGCCCTGGGGGGACAC	868
Db	1226	ACGTGAACCAACCATGCTGCTGTCGCCCAAGAAGAGGGAGGAAGAAGAGG-----	1280
QY	869	TGCCAGGGCCGCGCATGCTCATTTCTGAGCTCTGCTACTTTACAGGTCCTAACTGATAAAA	928
Db	1281	-AGTTGGTCCAGCCTACTACCGGCTTGAGCTGTGCTTCTGACTGGCTTAAGCGACACA	1339
QY	929	TGCGTAATGATTTTAAGCTGATGAAGACTTAGCCGTTTCATACAAGACTAACTCCAGAGC	988
Db	1340	TGCGCGCCGATTTCAACGTCATGAAGACATGGGCCAACACATCTGCTGTTGGACCCCAAG	1399
QY	989	AAAGGCAGCGTGAAGTGGGACGACTCATTCATATACATTCATAAAGCGATAATGTTCAA	1048
Db	1400	ACGATGTAGAACGCTCAGCGGCTTCATCAAAAGCTCAGCAGCAATGAAGAAGTGAAGA	1459
QY	1049	GGGAGCTTCGAGACTGGGGTTTGAGCTTTTGATTTCCAACTTACTGTCTCTCAGGAAGAA	1108
Db	1460	CTTACCTTAGATTCTCGGGCATGGAGTTGACAAAGAGCAGGTCMAACTGACTGGCGCTG	1519
QY	1109	TTTTGCAAAACAGAAAGATTCAACAAGTGGAAACAACTTTGATTAACAATCCACAATTG	1168
Db	1520	TCCTGCCACCTGAGAAACTATTCCAGAGGGGTAAACAGTTTCAGTTACAATCCGAGCAATG	1579
QY	1169	CAGATTGGTCCAAAGAACAGAGGTCGACCATTAATTAAGTGTGTTAAGCCACTAGATACT	1228
Db	1580	CGGACTGGTCCAGAGACACAGGGGTAAATGCCCTCACTGATGCAAAAGATCCTCAACA	1639
QY	1229	GGCTGTTGATCTTATACCGGAAGAAATATGAAGCAGGCCAATTCATTGATACAAAATCTAT	1288
Db	1640	GGAAGATCTTCTACACACCGCGGACCCACAGGGGCCAGGACTTCATCAAGTCACTCG	1699
QY	1289	TTAAAGTTACACAGCCATGGGCAATGCAAAATGAGAAAGCAATTAAGTATGAAGTG---G	1345
Db	1700	TAGAGTTCGCAAAACCCAATGGGTATGAACGTAAGCAGCAGAGATAGTGGAACTTCCCG	1759
QY	1346	ATGACAACTGAAGCCTACTTAAGTCTTTCAGCAAAAGGTCACAGCAGACACCCAGA	1405
Db	1760	ATGATCGTACAGAGACCTACAGAAGTCCCTCAGGCACAAATCGCACAGATACACAGA	1819
QY	1406	TAGTTGCTCTCTGTTTCAAGTAATCGGAAGGACAAATACGATGCTATTAAAAAATACC	1465
Db	1820	TTTGGTGGTGATCTTCGCAACGNAACAGAAAGGACCGGTACGATGCCATCAGAAAGCT	1879
QY	1466	TGTGTACAGATTGCCCTACCCCAAGTCAGTGTGTGTGGGCCGGAACCTTAGGCAAAACAGC	1525
Db	1880	CGGTGGTGAACCATCCATGCCCCAGCCAGGTCATCGTCTCCAGGACCCCTGTCACAGCAGC	1939
QY	1526	AAACTGTCATGGCCATGCTTCAAGATTGCCCTACAGATGAAGTGCAGATGGGAGGAG	1585
Db	1940	AGATGCTCATGTCGTGTGACCAACCAAGATTGCCATGTCAGATGAACTGCAAGATGGCGGAG	1999
QY	1586	AGCTCTGGAGGGTGGACATCCCCTGAAGCTCGTGATGATCGTTGGCATCGATTGTACC	1645
Db	2000	ACCTCTGGCAGTTCGAGATCCCGCTGAGCAACCTTGATGATCTCGGTATTGACAGTACC	2059

QY	1646	ATGACATGACAGCTGGCGGAGGTCAATCGCAGGATTTGTTGCCAGCATCAATGAAGGGA	1701
Db	2060	ATGACTCTCTGACCAAGGCGCTCCGCTCTCGGATTCGTGGCTTCATGAACAAGTCTC	2119
QY	1706	TGACCCGCTGTTCTTCAAGCTGCATATTTTCAGATAGAGGACGAGGAGCTGGTAGATGGC	1765
Db	2120	AAACCAAGTTTCTCTAGTTGTGCAATTCAGCACGCCACAGGAGAAATTCGGTGTAAAC	2179
QY	1766	TCAAAGTCTGCCGCAAGCGGCTCTGAGGCGTTTGGAAATAGCTGCAATGAGTATACGCCC	1825
Db	2180	TGAGTACCCCTCATGAACAATGCCCTGAAGCATACTATCAGATCAACGAGAGTTTCTCTG	2239
QY	1826	GCCGATCATCTGTATACCGGATGGCGTAGGAGACGCCGACGCTGAAAACTCGTGAACT	1885
Db	2240	AGCGTATCATCATCTTCCGTGATGTTGGTGACAGCCAAAGTCAATCTGGTCTTGACT	2299
QY	1886	ACGAAGTGGCACAGTTTTTGGGA--TTGTCTAAATCCATTGGTAGAGGTTTCAACCCCTA	1942
Db	2300	ATGAGCTCAAAACAATCAAGGACACCTTGGACAAAGTCTACCCCTCAAGGACGGTTCACA	2359
QY	1943	GACTAACGGTAAATTTGGTGTGAAGAAAAGAGTGAACACCAAGATTTTTTGTCTGCTGGAG	2002
Db	2360	AGCTGCGGTTGGTGTGTCGAAGAAGCGCATCAATAACCGCTCTTTTGGCAACTCTGAGAG	2419
QY	2003	GAAGACTTCAGAAATCCACTTCTCTGAAACAGTTTATTCATGTAGAGGTTTACAGACCCAGAT	2062
Db	2420	GAGGCGCTGCCAACCTCCCTCCCTGGAAACCGTTCATCGATGATGTCTGTCACCAAGCCCAAC	2479
QY	2063	GGTATACATTTTTTATTCGTGAGCGGCTGTGAGAAAGTGGTAGTGTCTCTCCACACATT	2122
Db	2480	TCTATGACTACTTTCATCATCAGTCAGAGTGTCCGAGGGCTCCGTGTCCCCACGTCCT	2539
QY	2123	ACAATCTCATCTATGACAAACAGCGGCTCTGAAGCCAGACACACATACAGCGCTTGACCTAC	2181
Db	2540	ACAACTTCTCTATGACACCACTGGTCTCAGCCAGACCATGACAGGTTGACCTAC	2598

RESULT 15

AB032605

LOCUS

AB032605

DEFINITION

Mus musculus mRNA for MILI, complete cds.

ACCESSION

AB032605

VERSION

AB032605.1

KEYWORDS

MILI (Miwi like).

SOURCE

Mus musculus (house mouse)

ORGANISM

Mus musculus

REFERENCE

1

AUTHORS

Kuramochi-Miyagawa,S., Kimura,T., Yomogida,K., Kuroiwa,A., Tadokoro,Y., Fujita,Y., Sato,M., Matsuda,Y. and Nakano,T.

TITLE

Two mouse piwi-related genes: miwi and mili

JOURNAL

Mech. Dev. 108 (1-2), 121-133 (2001)

MEDLINE

21463379

PUBMED

11578866

REFERENCE

2

AUTHORS

Miyagawa,S.K. and Nakano,T.

TITLE

Direct Submission

JOURNAL

Submitted (21-SEP-1999) Satomi K Miyagawa, Research Institute for Microbial Diseases, Department of Molecular Cell Biology, Yamadaoka 3-1, Suita, Osaka 565-0871, Japan (E-mail:smiyagawa@biken.osaka-u.ac.jp, Tel:81-6-6879-8362, Fax:81-6-6879-8362)

COMMENT

Sequence updated (04-Dec-1999).

FEATURES

Location/Qualifiers

1..4913

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/map="14D3"

/tissue_lib="testis"

1..4913

/gene="mili"

source

gene

2580	DB	AAAAAGTATTATGAGGTGAACCAATTGCTCCACAGAGAAAATTTGTGTGATCCAGAGATGGA	2639
1852	QY	GTAGGAGACGGCCAGCTGAAAAACACTGGTGAACTTACGAAGTGCCACAGTTTTTTTGGATTCT	1911
2640	DB	GTGCTGTGATGSCCAGCTAAAGACAGTTGCCAACTACGAGATCCCTCAGCTGCAGAGTGT	2699
1912	QY	CTAAAAATCCATTGGTAGAGGTTACAACCCTAGACTAAACGGTAATTTGTGTGAAGAAAAAGA	1971
2700	DB	TTTGAAGCCCTTTTGATAAC---TACCACCCCAAGATGGTGGTGTGTTGTAGTTTCAGAGAAA	2756
1972	QY	GTGAACACCAAGATTTTTTGCTCAGTCTGGAGGAAGACTTCAGAAATCCACTTCTCTGGAAACA	2031
2757	DB	ATCAGCACCAATCTGTACCTTTGCTGCTCTGTATCATCTTCGTAAACCCCTCCCCCGGACT	2816
2032	QY	GTTATTGATGTAGAGTTTACCAGACAGAAATGTFATGACTTTTTTATTCGTGAGCAGGCT	2091
2817	DB	GTGGTTGATCATACCATAAACAGCTGTGATGGGTGGATTTCTACGTTCTTGTGCCCATCAT	2876
2092	QY	GTGAGAAGTGGTAGTGTTTTCTCCACACATTAACAATGTCAATCTATCAACAACGCGCCCTG	2151
2877	DB	GTSCGACAGGCTGTGCATACCTACACACTACATCTGTGTTCTTGACACTGCAATCTG	2936
2152	QY	AAGCCAGACACATACAGCGCTTGACCTACAAGCTGTGCCACATCTATTATTAACCTGGCCA	2211
2937	DB	AGCCCTGATCACATGCAGAGTTGACTTTTCAAACTATGCCACATGTACTGGAAATTGGCCT	2996
2212	QY	GGTGTCAATTCGTGTTCTGTGCTCTTGGCAGTAGCCCAACAAGCTGCTTTTCTTGTGTGGC	2271
2997	DB	GGTACCATCCGAGTTCAGCTCCTTGCAAGTAGCCCAACAAGTAGCTTTCTGTGCCGA	3056
2272	QY	CAGAGTATTACAGAGACCAANCTGTCTACTGTCAAAACGGCTTTTACTACCTCTAA	2328
3057	DB	CAGATTTTGATCATGAGCCAGCCATCCAGCTGTGTGGAACTGTCTTCTCTGTAA	3113

Search completed: December 6, 2003, 14:28:52
Job time : 8522 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2003, 08:58:52 ; Search time 617 Seconds ;
(without alignments)

10185.238 Million cell updates/sec

Title: US-10-043-774B-1

Perfect score: 2328

Sequence: 1 atgattcttggtggaacac.....accgcctttactacctctaa 2328

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq 19Jun03:*

- 1: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 25: /SIDSL1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	ID	Description
1	2311	99.3	3472	21	Human piwi gene, d
2	1680.6	72.2	4064	21	Mouse piwi gene, d
3	1569	67.4	2292	24	Human colon cancer
4	458.4	19.7	2009	24	Human mRNA sequenc
5	395.4	17.0	423	21	Human breast and o
6	381	16.4	3649	22	cDNA sequence enco
7	365.4	15.7	367	25	Group III cDNA can
8	364	15.6	367	25	Group III cDNA can

9	330	14.2	2272	22	AAH15360	Human cDNA sequenc
10	310.6	13.3	2838	23	ABL10571	Drosophila melanog
11	286.4	12.3	300	20	AZ13404	Human gene express
12	263.4	11.3	3524	23	ABL10535	Drosophila melanog
13	256.6	11.0	3047	21	AAH07586	Drosophila piwi ge
14	237.2	10.2	1733	22	AAH14032	Human cDNA sequenc
15	207.4	8.9	209	25	ABZ19102	Group III cDNA can
16	184	7.9	5671	23	ABL10570	Drosophila melanog
17	171.8	7.4	548	24	ABN62610	Human cancer relat
18	149.2	6.4	1026	23	ABN70144	DNA encoding novel
19	138	5.9	7667	23	ABL10534	Drosophila melanog
20	113.2	4.9	786	22	AAH05561	Human cDNA clone
21	111.2	4.8	2914	22	AAK53425	Human polynucleoti
22	111.2	4.8	2914	25	ABX13625	Human cytokine or
23	111.2	4.8	3050	22	AAH16058	Human cDNA sequenc
24	111.2	4.8	3736	25	ABX34752	Human mddt cDNA SE
25	104.2	4.5	325	22	AAH84876	Human polynucleoti
26	95.6	4.1	444	23	AAH57327	cDNA #3 encoding p
27	91.2	3.9	601	23	AAH77101	DNA encoding novel
28	90	3.9	646	23	AAH77101	DNA encoding novel
29	86.6	3.7	3996	22	AAH14510	Human cDNA sequenc
30	83.2	3.6	2951	22	AAH24562	Translation initia
31	82	3.5	642	22	AAH07789	Human cDNA clone
32	82	3.5	1968	22	AAH14966	Human cDNA sequenc
33	81.8	3.5	3011	24	ABQ99304	Human coding sequ
34	81.8	3.5	3580	24	ABS78717	Human cDNA encodin
35	79.8	3.4	516	22	AAH05278	Human cDNA clone
36	79	3.4	734	20	AAZ15059	Human gene express
37	78.8	3.4	771	22	AAK58246	Human immune/haema
38	78.6	3.4	213	24	ABS17735	Human genome-deriv
39	69.6	3.0	1158	21	AAH50998	Arabidopsis thalia
40	69.6	3.0	3195	21	AAH50992	Arabidopsis thalia
41	69.6	3.0	3336	21	AAH47079	Arabidopsis thalia
42	68	2.9	364	22	ABA56879	Human foetal liver
43	68	2.9	384	22	ABA26498	Probe #4964 for ge
44	68	2.9	364	24	ABS05149	Human genome-deriv
45	64.2	2.8	166	22	ABA69444	Human foetal liver

ALIGNMENTS

RESULT 1

AAA07588

ID AAA07588 standard; DNA; 3472 BP.

XX AAA07588;

AC AAA07588;

XX 29-AUG-2000 (first entry)

XX Human piwi gene, designated hiwi.

XX Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;

XX anaemia; immunodeficiency; male infertility; human; ds.

XX Homo sapiens.

XX Key

XX Location/Qualifiers

XX CDS

XX 167..2752

XX /*tag= a

XX /product= hiwi

XX /transl_except= (pos:392..394; aa:Xaa)

XX /transl_except= (pos:1073..1075; aa:Xaa)

XX /transl_except= (pos:2369..2370; aa:Xaa)

XX /note= "Xaa= leu or ile"

XX WO200032039-A1.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX (UYDU-) UNIV DUKE.
PA Lin H;
XX WPI: 2000-412085/35.
DR P-PSDB; AAY90235.
XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
PT therapy of diseases such as cancer and in various research and
PT diagnostic applications -
XX Claim 19; Page 189-194; 201pp; English.
XX This sequence encodes the human piwi family protein, designated
CC hiwi. The piwi family nucleic acids and polypeptides are used in gene
CC therapy of diseases such as cancer and also in various research and
CC diagnostic applications. The sequences can also be used to treat
CC tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX Sequence 3472 BP; 1044 A; 694 C; 796 G; 935 T; 3 other;
SQ
Query Match 99.3%; Score 2311; DB 21; Length 3472;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 2314; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 8 TTGGTGTGAACACAGCGCAGACTAGACCATGTTAAAGATCAAAAACAGGTTCTTCAG 67
DB 432 TTGGTGTGAATCAAGCGCAGAACCTAGACCATGTTAAAGATCAAAAACAGGTTCTTCAG 491
QY 68 GCATTATAGTAGGTTAAGCACTAACCACTTCGGGTGACATCCCGTCCCGAGTGGGCT 127
DB 432 GCATTATAGTAGGTTAAGCACTAACCACTTCGGGTGACATCCCGTCCCGAGTGGGCT 551
QY 128 TATATCAGTATCACATGACTATAACCCACTGATGAAGCCAGAGACTCCGTTGAGCTC 187
DB 552 TATATCAGTATCACATGACTATAACCCACTGATGAAGCCAGAGACTCCGTTGAGCTC 611
QY 188 TTCCTTTTCAACACAGAGATCTAATGGAAGTGTGCTTTGATGGAACAGATATTAT 247
DB 612 TTCCTTTTCAACACAGAGATCTAATGGAAGTGTGCTTTGATGGAACAGATATTAT 671
QY 248 TTTTACCTTAAAGAGCTACAGCAAAAGTTTACTGAAAGTTTATAGTAAGCCCGGAATGGAG 307
DB 672 TTTTACCTTAAAGAGCTACAGCAAAAGTTTACTGAAAGTTTATAGTAAGCCCGGAATGGAG 731
QY 308 AGGATGTGAGGATAACGATCACTTTAAACAAATGAACCTTCCACTATACCACTTGT 367
DB 732 AGGATGTGAGGATAACGATCACTTTAAACAAATGAACCTTCCACTATACCACTTGT 791
QY 368 TGCAGTCTATATATATTTTCAGAGGCTTTTGAAGATCATGAATTTGCAACAAATG 427
DB 792 TGCAGTCTATATATATTTTCAGAGGCTTTTGAAGATCATGAATTTGCAACAAATG 851
QY 428 GACGAAATATTATTAACCCAAATGACCAATGATATTCGAAGTCAACAGTTGGTGATTT 487
DB 852 GACGAAATATTATTAACCCAAATGACCAATGATATTCGAAGTCAACAGTTGGTGATTT 911
QY 488 GGCCTGGCTTCACTACTTCCACTTTCAGTATGAAGAACAGCATCATGCTCTGCACTGAG 547
DB 912 GGCCTGGCTTCACTACTTCCACTTTCAGTATGAAGAACAGCATCATGCTCTGCACTGAG 971
QY 548 TTAGCCATTAAGTCTTCCAGTGAAGCTGTTTGGATTTTCATGTTCACTTTTATATATC 607
DB 972 TTAGCCATTAAGTCTTCCAGTGAAGCTGTTTGGATTTTCATGTTCACTTTTATATC 1031
QY 608 AGACAGAGAAACATAAATTTCAAGAACAGTTTCCAAAGAACTAATAGTTTGTGTTTC 667
DB 1032 AGACAGAGAAACATAAATTTCAAGAACAGTTTCCAAAGAACTAATAGTTTGTGTTTC 1091
QY 668 TTACCAAGTATTAACATTAAGACATACAGAGTGGATGATATTGACTGGGACAGAAATCCA 727
DB 1092 TTACCAAGTATTAACATTAAGACATACAGAGTGGATGATATTGACTGGGACAGAAATCCA 1151

QY 728 AGAGCACTTTTAAAGAAAGCCGCTCTGAAGTCAAGCTTCTTAGAATACATACAGAAAGC 787
DB 1152 AGAGCACTTTTAAAGAAAGCCGCTCTGAAGTCAAGCTTCTTAGAATACATACAGAAAGC 1211
QY 788 AATACACACAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCAGCCCAAGAGAA 847
DB 1212 AATACACACAGAGATCACCGACTTGAAGCAGCCTGTCTTGGTCAGCAGCCCAAGAGAA 1271
QY 848 GGCCTGGCTTCACTACTTCCACTTTCAGTATGAAGAACAGCATCATGCTCTGCACTGAG 907
DB 1272 GGCCTGGCTTCACTACTTCCACTTTCAGTATGAAGAACAGCATCATGCTCTGCACTGAG 1331
QY 908 TTAAGTCTTAACCTGATAAATGCGTAATGATTTTAAAGTGAATGAAAGCTTACGCTTC 967
DB 1332 TTAAGTCTTAACCTGATAAATGCGTAATGATTTTAAAGTGAATGAAAGCTTACGCTTC 1391
QY 968 ATACAGACTTAACCTCCAGAGCAAGGCGAGCTGAGTGGGACGACTCATGATGATGATTC 1027
DB 1392 ATACAGACTTAACCTCCAGAGCAAGGCGAGCTGAGTGGGACGACTCATGATGATGATTC 1451
QY 1028 ATAAAAACGATAATGTTTCAAGGAGGAGCTTCCAGACTGGGGTTTTCAGCTTTTCACTTCC 1087
DB 1452 ATAAAAACGATAATGTTTCAAGGAGGAGCTTCCAGACTGGGGTTTTCAGCTTTTCACTTCC 1511
QY 1088 TACTGTCTTCTCAGGAGAAATTTTGCAGAGAGAGATTTGCAAAACAGAAAGATTTCAACAAAG 1147
DB 1512 TACTGTCTTCTCAGGAGAAATTTTGCAGAGAGAGATTTGCAAAACAGAAAGATTTCAACAAAG 1571
QY 1148 TTGATTAACATCCCAATTTGCGATTTGCTTGAATGCTATAGCGAAGAAATATGAAGCAGCA 1207
DB 1572 TTGATTAACATCCCAATTTGCGATTTGCTTGAATGCTATAGCGAAGAAATATGAAGCAGCA 1631
QY 1208 GTGTTAAAGCCACTAGATAAATGCTGTTGATGCTATAGCGAAGAAATATGAAGCAGCA 1267
DB 1632 GTGTTAAAGCCACTAGATAAATGCTGTTGATGCTATAGCGAAGAAATATGAAGCAGCA 1691
QY 1268 ATTCAATTGATACAAAATCTATTAAAGTTAAAGTTAAAGTTAAAGTTAAAGTTAAAGTTAA 1327
DB 1692 ATTCAATTGATACAAAATCTATTAAAGTTAAAGTTAAAGTTAAAGTTAAAGTTAAAGTTAA 1751
QY 1328 CAATAATGATTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1387
DB 1752 CAATAATGATTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1811
QY 1388 TCACAGCAGACACCCAGATAGTTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1447
DB 1812 TCACAGCAGACACCCAGATAGTTGCTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1871
QY 1448 ATGCTATTAATAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAAGTGTGTGTGTGTGTGT 1507
DB 1872 ATGCTATTAATAAATACCTGTGTACAGATTGCCCTACCCCAAGTCAAGTGTGTGTGTGTGTGT 1931
QY 1508 GAACTTGAAGCAACACAGCAACTGCTCATGGCCATTGCTTACAAAGATTGCCCTACAGATGA 1567
DB 1932 GAACTTGAAGCAACACAGCAACTGCTCATGGCCATTGCTTACAAAGATTGCCCTACAGATGA 1991
QY 1568 ACTCAAGATGGGAGGAGAGCTCTGAGGGTGGACATCCCCCTCAAGCTCGTGATGATCG 1627
DB 1992 ACTCAAGATGGGAGGAGAGCTCTGAGGGTGGACATCCCCCTCAAGCTCGTGATGATCG 2051
QY 1628 TTGCGATCGATTGTTTACATGACATGACAGCTGGGGGAGGCTCAATCGAGGATTTGTTG 1687
DB 2052 TTGCGATCGATTGTTTACATGACATGACAGCTGGGGGAGGCTCAATCGAGGATTTGTTG 2111
QY 1688 CCAGCATCAATGAAGGATGACCCGCTGGTCTCAAGCTGATATTTTCAAGATGAGGAC 1747
DB 2112 CCAGCATCAATGAAGGATGACCCGCTGGTCTCAAGCTGATATTTTCAAGATGAGGAC 2171
QY 1748 AGGAGCTGGTAGATGGGCTCAAGCTGCGCTGCAAGCGGCTCTGAGGGCTTGGATAGCT 1807
DB 2172 AGGAGCTGGTAGATGGGCTCAAGCTGCGCTGCAAGCGGCTCTGAGGGCTTGGATAGCT 2231

CC The invention relates to a human colon tumour expressed polynucleotide
 CC (I) encoding a polypeptide (II, ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8689-ABV8289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i), or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (ii) and (iii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 2292 BP; 692 A; 435 C; 512 G; 653 T; 0 other;

Query Match 67.4%; Score 1569; DB 24; Length 2292;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1569; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 760 GTCAGCTTCTTAGAATCTACAGGAAGCAATACACCAAGAGATCACCGACTTGAAGCAG 819
 DB 4 GTCAGCTTCTTAGAATCTACAGGAAGCAATACACCAAGAGATCACCGACTTGAAGCAG 63
 QY 820 CCTGTCTTGGTCAGCCAGCCCAAGAGAGGGGGCCCTGGGGGACACTGCCAGGGCCT 879
 DB 64 CCGTCTTGGTCAGCCAGCCCAAGAGAGGGGGCCCTGGGGGACACTGCCAGGGCCT 123
 QY 880 GCCATGCTCATCTCGAGCTCTGTATCTTACAGTCTTAACGTAAATAATGGTAAATGAT 939
 DB 124 GCCATGCTCATCTCGAGCTCTGTATCTTACAGTCTTAACGTAAATAATGGTAAATGAT 183
 QY 940 TTTAAGTGATGAAGACTTACGCTTCATACAGACTTAACCTCCAGAGCAAGGCGAGCT 999
 DB 184 TTTAAGTGATGAAGACTTACGCTTCATACAGACTTAACCTCCAGAGCAAGGCGAGCT 243
 QY 1000 GAACTGGGACGACTCATTGATTACATTTATATAAAACGATAATGTTCAAGGGAGCTTGA 1059
 DB 244 GAAGTGGGAGGACTCATTGATTACATTTATATAAAACGATAATGTTCAAGGGAGCTTGA 303
 QY 1060 GACTGGGTTGAGCTTGTGANTTCAACTTACTGCTCTTCTCAGGAAGAAATTTTGAACA 1119
 DB 304 GACTGGGTTGAGCTTGTGANTTCAACTTACTGCTCTTCTCAGGAAGAAATTTTGAACA 363
 QY 1120 GAAAGATTCACCAAGTGGGAAACATTTGATTAACATCCCAATTTGCGAGTTGCTCC 1179
 DB 364 GAAAGATTCACCAAGTGGGAAACATTTGATTAACATCCCAATTTGCGAGTTGCTCC 423
 QY 1180 AAAGAAACAAGAGTGCACCATTAATTAGTGTAAAGCCACTAGATAAAGCTGTTGATC 1239
 DB 424 AAAGAAACAAGAGTGCACCATTAATTAGTGTAAAGCCACTAGATAAAGCTGTTGATC 483
 QY 1240 TATACGGAAGAATAATGAAGCAGCAATTCATTGATACAAAATCTATTAAAGTTACA 1299
 DB 484 TATACGGAAGAATAATGAAGCAGCAATTCATTGATACAAAATCTATTAAAGTTACA 543
 QY 1300 CCAGCATGGGCAATCAATGAGAAAGCAATTAATGATTGAAGTGTGATGATGATGATG 1359
 DB 544 CCAGCATGGGCAATCAATGAGAAAGCAATTAATGATTGAAGTGTGATGATGATGATG 603
 QY 1360 GCCTACTTAAGAGTCTTACAGCAAAAGGTTCACAGCAGACACCCAGATAGTGTCTGTCTG 1419
 DB 604 GCCTACTTAAGAGTCTTACAGCAAAAGGTTCACAGCAGACACCCAGATAGTGTCTGTCTG 663
 QY 1420 TTGTCAAGTAATCGGAAGCAAAATACGATGCTATTAAATAATCTGTGTACAGATTGC 1479
 DB 664 TTGTCAAGTAATCGGAAGCAAAATACGATGCTATTAAATAATCTGTGTACAGATTGC 723
 QY 1480 CTTACCCCAAGTCAGTGTGTGTGGCCGAACTTAGGCAAAACAGCAAACTGTCTGAGCC 1539
 DB 724 CTTACCCCAAGTCAGTGTGTGTGGCCGAACTTAGGCAAAACAGCAAACTGTCTGAGCC 783
 QY 1540 ATTGCTACAAAGATTGCCCTACAGATGAATGCAAGATGGGAGGAGAGCTCTGGAGGGTG 1599

DB 784 ATTGCTACAAAGATTGCCCTACAGATGAATGCAAGATGGAGGAGAGCTCTGGAGGGTG 843
 QY 1600 GACATCCCTCGAAGCTCGTATGATCGTTGGGCAATGATTTTACCATGATGACAGCT 1659
 DB 844 GACATCCCTCGAAGCTCGTATGATCGTTGGGCAATGATTTTACCATGATGACAGCT 903
 QY 1660 GGGGGAGGTCAATCGCAGGATTTGTTCCAGCATCAATGAAGGATGACCCGCTGGTTC 1719
 DB 904 GGGGGAGGTCAATCGCAGGATTTGTTCCAGCATCAATGAAGGATGACCCGCTGGTTC 963
 QY 1720 TCACGCTGCATATTTCAGGATAGAGGACAGAGCTGTGTAGTGGGCTCAAAAGTCTGCTG 1779
 DB 964 TCACGCTGCATATTTCAGGATAGAGGACAGAGCTGTGTAGTGGGCTCAAAAGTCTGCTG 1023
 QY 1780 CAACGGCTCTGAGGGCTTGGATAGCTGCATAGTACATGCCAGCCGATCATGCTG 1839
 DB 1024 CAACGGCTCTGAGGGCTTGGATAGCTGCATAGTACATGCCAGCCGATCATGCTG 1083
 QY 1840 TACCGCATGGCTGAGGACGGCCAGCTGAAACACTGGTGAACTACGAAGTGCCACAG 1899
 DB 1084 TACCGCATGGCTGAGGACGGCCAGCTGAAACACTGGTGAACTACGAAGTGCCACAG 1143
 QY 1900 TTTTGGATTGTCTAAATCCATTGTTAGAGGTTTACAAACCTAGACTAACCGTAAATGCTG 1959
 DB 1144 TTTTGGATTGTCTAAATCCATTGTTAGAGGTTTACAAACCTAGACTAACCGTAAATGCTG 1203
 QY 1960 GTGAAGAAAGAGTGAACACACAGATTTTGTCTCAGTCTGAGGAGAGCTTCAGATCCA 2019
 DB 1204 GTGAAGAAAGAGTGAACACACAGATTTTGTCTCAGTCTGAGGAGAGCTTCAGATCCA 1263
 QY 2020 CTTCTCTGAAACAGTATTGATGTAGAGTTACCAAGACAGAGTACGATGATGATGATGATG 2079
 DB 1264 CTTCTCTGAAACAGTATTGATGTAGAGTTACCAAGACAGAGTACGATGATGATGATGATG 1323
 QY 2080 GTGAGCCAGGCTGTGAGAGTGTGTGTCTCTCCACACATTAATGATCATCTATGAC 2139
 DB 1324 GTGAGCCAGGCTGTGAGAGTGTGTGTCTCTCCACACATTAATGATCATCTATGAC 1383
 QY 2140 AACAGCGGCTTGAAGCAGACACATACAGGCTTGACCTACAGAGCTTGCCACATCTAT 2199
 DB 1384 AACAGCGGCTTGAAGCAGACACATACAGGCTTGACCTACAGAGCTTGCCACATCTAT 1443
 QY 2200 TACAAGTGGCAGGTGTCATTGTTCTGCTCTGCTCCAGTACGCCACAGAGTGGCT 2259
 DB 1444 TACAAGTGGCAGGTGTCATTGTTCTGCTCTGCTCCAGTACGCCACAGAGTGGCT 1503
 QY 2260 TTTCTTGTGGCCAGAGTATTACAGAGAGCCAAATCTGTCACTGTCAAAACCGCTTTAC 2319
 DB 1504 TTTCTTGTGGCCAGAGTATTACAGAGAGCCAAATCTGTCACTGTCAAAACCGCTTTAC 1563
 QY 2320 TACCTCTAA 2328
 DB 1564 TACCTCTAA 1572

RESULT 4

ABQ61109
 ID ABQ61109 standard; cdna; 2009 BP.

XX AC ABQ61109;

XX XX 26-FEB-2003 (first entry)

XX DE Human mRNA sequence encoding sequence.

XX KW Neuroprotective; immunomodulator; cancer;

XX KW cytosolic; anti-inflammatory; gene therapy; nutritional supplement;

XX KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;

XX KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;

XX KW vulnary; gene; ss.

XX OS Homo sapiens.

XX WO200231111-A2.
 PN
 XX
 XX PD 18-APR-2002.
 XX
 XX 11-OCT-2001; 2001WO-US27760.
 XX
 XX 12-OCT-2000; 2000US-0687527.
 PR
 XX (HYSE-) HYSEQ INC.
 XX
 XX PA Xue YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Tang AJ, Yang Y, Wehrman T, Drmanac RT;
 BL
 XX
 XX DR WPI; 2002-426278/45.
 DR N-ESDB; ABP43865.
 XX
 XX New polypeptides and their encoded proteins, useful as nutritional
 PT sources of supplements, or in gene therapy, particularly for treating
 PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
 PT inflammation -
 XX
 XX Claim 1; SEQ ID # 322; 357pp + sequence listing; English.
 XX
 XX The invention relates to 446 newly isolated polynucleotide sequences.
 CC The activity of polynucleotides of the invention may be described as,
 CC anti-inflammatory, neuroprotective, immunomodulator, cytostatic and
 CC anti-inflammatory. Compositions comprising nucleic acids of the invention
 CC are useful for treating a mammalian subject, or as nutritional sources or
 CC supplements. These are useful in gene therapy, particularly for treating
 CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
 CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
 CC inflammation. The nucleic acids and polypeptides are also useful in
 CC diagnostic and research methods. The sequences given in records
 CC AB060788-AB061233 represent polynucleotides of the invention.
 CC NOTE: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.
 XX
 XX SQ Sequence 2009 BP; 595 A; 411 C; 458 G; 545 T; 0 other;
 Query Match 19.7%; Score 458.4; DB 24; Length 2009;
 Best Local Similarity 57.2%; Pred. No. 6.9e-126;
 Matches 894; Conservative 0; Mismatches 661; Indels 9; Gaps 3;

Qy 765 CTTCTTGAATACACAGAGCAATACCAACAGAGATCCCGACTTGAAGCAGCCTGT 824
 Db 15 CTAGGTGATTACTACAAGCAGCAGTATGATATTACTGTATCGGACCTGATCAGCCAT 74
 Qy 825 CTTGTCTAGCCAGCCCAAGAGAGCGGGCCCTGGGGGGACATGCCAGGCGCTGCCAT 884
 Db 75 CTTGTGTAGTCTGTTAAG---AAGAGAGAAATGACACACAGTGAAGCTCGGCCA 131
 Qy 885 GCTCATCTCCTGAGCTCTCTATCTTACAGGTCTTAATGATATAAATGCGTAATGATTTAA 944
 Db 132 CTTGATACCTTGAGCTCTCTCTTCTTAAAGGCTGACTGACAGGCAACATCTGATTTCA 191
 Qy 945 CGTGATGAAGACTTAGCCGTTTACACAGACTAACTCCAGAGCAAAAGCGAGCGTGAAGT 1004
 Db 192 GCTGATGAAGGCTGTGGCTGAAAAGACACGCTCTCAGTCTTTAGCCCGCAGCAGCGCT 251
 Qy 1005 GGGAGCTCATTTGATTAATTCATTAATAAAGATATGTTCAAGGGAGCTTCGAGACTG 1064
 Db 252 GCCCAGGCTTGTGGACAAATCCAGAGAAATACCAATGCTCGCTTGAACATAGAGACCTG 311
 Qy 1065 GGGTTTGAGCTTTGATTCACAACTTACTCTCTCTCAGAGAAATTTTGCAGAAAGAAA 1124
 Db 312 GGGACTGATTTTGAAGCCAGATA---TCTCTGACTGGCGGATGTGCTTTCAGAAAA 368
 Qy 1125 GATTCACCAAGGTGGAAGAAACATTTGATTAACAATCCAAATTTGAGATTTGGTCCAAAGA 1184
 Db 369 AATATTATGCAAGACCAACATAT---GTCAACCTGTGTCTGCTGACTGGTCCAGGA 425

Qy 1185 AACAGAGGTGCACCATTATTAAGTGTAAAGCCACTAGATAAATGGCTGTGTATCTATAC 1244
 Db 426 TATTTCGAACCTTGCAGAGATTTTAAATGCACAGTCTTTTGAATACCTGGTTGATTTATGTAG 485
 Qy 1245 GCGAAGAAATATGAAGCAGCCCAATTCATTGATACAAAATCTATTTAAAGTTACACCCAGC 1304
 Db 486 CGACAGAACTGAAATATGTTCCGAGAGCTTTTGAATGCTTGAAGAGAGTTGAGGTTTC 545
 Qy 1305 CATGGGCATGCAAAATGAGAAAAGCAATATGATGAGTGAAGTGGATGACAGAACTGAAGCCTA 1364
 Db 546 CATGGGATTTAAATGCTGGACTACCCCAAATCATAAAGTACAGAAATCCAGCTGCATT 605
 Qy 1365 CTTAAGAGCTTTACAGCAAAAGGTACAGCAGACACCCAGATAGTTGCTGTCTGTCTGTC 1424
 Db 606 TGTTAGAGCTTATACAGCAATATGTTGATCTCTGATGTTTCACTGCTGTAATGTGCTATTCGCC 665
 Qy 1425 AAGTAATCGAAGGACAAATACGATGCTATTAAAAAATACCTGTGTACAGATTCGCCCTAC 1484
 Db 666 TTCTAATCAGAGACCTATTATGATTCATTAAAAAATATTTGAGCTCAGATGCCAGT 725
 Qy 1485 CCCAAGTCACTGTGTGGTGGCCGAACTTTAGGCAAAACAGCAAACTGTCTATGGCCATTGC 1544
 Db 726 CCCAAGCCAATGTGTCTTCTCGGACCTTGAATAACAGGGCATGATGATGATGATCGC 785
 Qy 1545 TACAAGATTTGCCCTACAGATGAATCTCAAGATGGAGGAGAGCTCTGGAGGTGGACAT 1604
 Db 786 CACCAAGATCGCTATGAGATGACTTTCAAGCTCGGAGCGAGCTGTGGGTGTGGAAAT 845
 Qy 1605 CCCCTCAAGCTCGTGTATGATGCTGTGCATCGATGTTTACCATGACATGACAGCTGGGCG 1664
 Db 846 ACCTTAAATCCCTGATGTTGCTGGTATGATGCTGTAAGATGCATCAGCAAGGA 905
 Qy 1665 GAGTCAATCGCAGGATTTTGTCCAGCATCAATGAAGGATGACCCGCTGGTTCTCACG 1724
 Db 906 CGTGATGTTGTTGGATGCTGGCGAGTGTAAACCCAGAAATCACAGGTGGTTTCCCG 965
 Qy 1725 CTGCATTTTCAGATAGAGGACAGGAGCTGTGTAGTGGGCTCAAAGTCTGCCTGCAGC 1784
 Db 966 CTGTATCTTTACAGAGAAACAATGATGTTGTCAGATTTGCTTGAAGCTTTTCATGACTGG 1025
 Qy 1785 GGCTCTGAGGCTTGGAAATAGCTGCAATGATGATCCCGCCAGCGGATCATCGTGTACCG 1844
 Db 1026 AGCACTCAACAATGTPACAGTACAAATCATGATTTCCAGACCGGATTAATGTTGATCCG 1085
 Qy 1845 CGATGGCTTAGGAGACGGCCAGCTGAAAAACATGGTGAACACTACGAAGTGCACAGTTTTT 1904
 Db 1086 TGCTGGTGTAGGGGATGGTCAAGCTGAAAAACATTTATGAAATATGAAGTCCACAGCTGCT 1145
 Qy 1905 GGATTGTCTAAATCCATTGGTAGAGTTTACAACCTTAGACTAACCGTAAATGTTGGTGAA 1964
 Db 1146 GAGCAGTGTGGCAGAAATCCAGCTCAAAATACAGCTCAAGACTGTCCGTGATTTGGTCCAG 1205
 Qy 1965 GAAAAGAGTGAACACACAGATTTTGTCTGCTGAGTGGAGGAAAGCTTCAGAAATCCACTTCC 2024
 Db 1206 GAAGAAGTGCATGCCAGATCTTTTACCGAATGAACCGCAGCTGTACAGAACCCCACT 1265
 Qy 2025 TGAACAGTATTGATGATGAGGTTTACAGACCAAGAAATGATGATGATTTTATCTGTGAG 2084
 Db 1266 TGGCACTGTTGTGGATTCAGAAAGCAACACGTAACGAATGGTATGACTTTTATCTGATCAG 1325
 Qy 2085 CCAGGCTGTGAGAGTGTGTGTTTCTCCCAACATTAACAATGCTCATCTATGACACAG 2144
 Db 1326 CCAGGTGGCTTCCGGGAACTGTGTAGTCTTACTCTACTATAATGTCACTATGATGACAA 1385
 Qy 2145 CGGCCTGAAGCCAGACACACATACAGCGCTTGACCTCAAGCTGTGCCACATCTATTACAA 2204
 Db 1386 CGGCTTGAAGCCGACCATATGAGAGACTTACATTCAAATTTGGCCACTGTACTACAA 1445
 Qy 2205 CTGSCCAGGTTGTCATTCGTGTTCTCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2264
 Db 1446 CTGSCCGGCGATAGTCAGTGTCCAGCAACCATGTCAGTATGCTCAGCAAGCTGACCTTTCT 1505
 Qy 2265 TGTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCTACTGTCTCTCTCTCTCTCTCTCTCT 2324

Db 1506 GGGGGCACAAGCATTTCATTAAGAACCCAGCTGCTGGAATTAGCCCAACCATCTCTTCTACCT 1565
 QY 2325 CTAA 2328
 Db 1566 GTGA 1569

RESULT 5
 AAF21880
 ID AAF21880 standard; DNA; 423 BP.
 XX
 XX AAF21880;
 XX 27-MAR-2001 (first entry)
 XX Human breast and ovarian cancer associated antigen gene SEQ ID 267.
 DE Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
 KW neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 KW antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;
 KW Addison's disease; allergy; autoimmune haemolytic anaemia;
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
 KW cardiovascular disorder; wound healing; neurological disease; ds.
 XX Homo sapiens.
 XX WO200055173-A1.
 XX 21-SEP-2000.
 XX 08-MAR-2000; 2000WO-US05881.
 XX 12-MAR-1999; 99US-0124270.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Ruben SM;
 XX WPI; 2000-611515/58.
 XX P-PSDB; AAB58977.
 XX New human breast and ovarian cancer associated gene sequences and the
 PT polypeptides encoded by these genes, useful in the prevention,
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular
 PT disorders and neurological diseases -
 XX
 PS Claim 1; Page 686; 1299pp; English.
 CC Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are
 CC associated with breast and ovarian cancer. Included in the invention are
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the
 CC isolation and characterisation of the DNA and protein sequences of the
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;
 CC neotropic; neuroprotective; antiviral; antiallergic; hepatotropic;
 CC antidiabetic; antiinflammatory; antitumor; vulnery; anticonvulsant;
 CC antibacterial; antifungal; antiparasitic and cardiant activity. The
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,
 CC particularly breast and ovarian cancer. The nucleic acid sequences,
 CC proteins, agonists and antagonists may also be used in the diagnosis,
 CC prevention and treatment of immune disorders e.g. Addison's disease,
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid
 CC arthritis and ulcerative colitis; cardiovascular disorders such as
 CC myocardial ischaemias; wound healing; neurological diseases such as
 CC cerebral anoxia and epilepsy; and infectious diseases.
 XX
 SQ Sequence 423 BP; 104 A; 98 C; 129 G; 88 T; 4 other;

Query Match 17.0%; Score 395.4; DB 21; Length 423;
 Best Local Similarity 98.8%; Pred. No. 1.8e-107;
 Matches 404; Conservative 4; Mismatches 0; Indels 1; Gaps 1;
 QY 1474 GATTGCCCTACCCCAAGTCAAGTGTGGTGGCCGCAACCTTAGGCAACAGCAACTGTG 1533
 Db 14 GATTGCCCTACCCCAAGTCAAGTGTGGTGGCCGCAACCTTAGGCAACAGCAACTGTG 73
 QY 1534 ATGGCCATTGCTACAAAGATTGCCCTACAGATCAACTGCAAGATGGGAGGAGACTCTGG 1593
 Db 74 ATGGCCATTGCTACAAAGATTGCCCTACAGATCAACTGCAAGATGGGAGGAGACTCTGG 133
 QY 1594 AGGGTGGACATCCCCCTGAAGCTCGTGCATCATCGTTGGCATCGATTGTTACCATGACATG 1653
 Db 134 AGGGTGGACATCCCCCTGAAGCTCGTGCATCATCGTTGGCATCGATTGTTACCATGACATG 193
 QY 1654 ACAGCTGGGGGAGGAGTCAATCGCAGGATTTGTCAGCATCAATGAAGGAGTACCCGC 1713
 Db 194 ACAGCTGGGGGAGGAGTCAATCGCAGGATTTGTCAGCATCAATGAAGGAGTACCCGC 253
 QY 1714 TGCTTCTCAGCTGTCATATTTTCAGGATAGAGACAGAGCTGGTAGATGGCTCAAAAGTC 1773
 Db 254 TGCTTCTCAGCTGTCATATTTTCAGGATAGAGACAGAGCTGGTAGATGGCTCAAAAGTC 313
 QY 1774 TGCTTCAAGCGGCTCTGAGGGCTTGCAATAGCTGCAATGAGTACATGCCAGCCGATC 1833
 Db 314 TGCTTCAAGCGGCTCTGAGGGCTTGCAATAGCTGCAATGAGTACATGCCAGCCGATC 373
 QY 1834 ATCGTGTACCCGATGGCTGAGAGACGGCCAGCTGAAACACTGGTGA 1882
 Db 374 ATCGTGTACCCGCTGAGGAGACGGCCAGCTGAAACACTGGTGA 421

RESULT 6
 AAS13630
 ID AAS13630 standard; cDNA; 3649 BP.
 XX
 XX AAS13630;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE cDNA sequence encoding mammalian Spg16.
 XX
 KW Mammalian; reproductive-specific protein; male infertility; gene therapy;
 KW spermatogenesis; sperm count disorder; anti infertility;
 KW reproduction; ss.
 XX
 OS Mammalia.
 XX
 PN WO200166752-A2.
 XX
 PD 13-SEP-2001.
 XX
 PF 07-MAR-2001; 2001WO-US07371.
 XX
 PR 07-MAR-2000; 2000US-0187518.
 PR 12-JAN-2001; 2001US-0261557.
 XX
 XX (WHED) WHITEHEAD INST BIOMEDICAL RES.
 PA
 XX Wang PJ, Page DC;
 XX WPI; 2001-570774/64.
 XX P-PSDB; AAU07866.
 DR
 XX Novel reproduction-specific protein, useful for treating disorders of
 XX reduced sperm count, enhancing/increasing sperm count and/or sperm
 XX activity -
 XX
 PS Claim 1; Fig 15; 151pp; English.
 XX
 CC The present invention relates to the isolation of novel mammalian and
 CC human reproductive-specific proteins (AAU07859-AAU07899), and the

CC nucleic acids encoding them. The nucleic acids encoding
 CC reproductive-specific proteins are useful for diagnosing infertility
 CC which is a result of reduced sperm count, reduced sperm motility,
 CC malformed sperm or combinations of these. The sequences of the invention
 CC are useful as markers for spermatogonial cells, for identifying genes or
 CC proteins characteristic of male infertility, diagnosing or aiding in
 CC the diagnosis of infertility in men, and for contraception in which
 CC sperm production or sperm count is reduced or defective sperm is
 CC produced. Antibodies to reproductive-specific proteins are useful for
 CC determining the presence of these proteins in a sample obtained from a
 CC man being assessed for infertility, for identifying the expression of
 CC genes in particular cell type or particular developmental stage, for
 CC studies of spermatogenesis, and for immunofluorescence of germ cells or
 CC in western blots for assessing the presence of the protein the antibody
 CC binds. The sequences of the invention are also useful for treating
 CC disorders of reduced sperm count, and for increasing sperm count and/or
 CC sperm activity. The nucleic acids of the invention are useful in gene
 CC therapy. AAS13623-AAS13647 represent cDNA sequences encoding for the
 CC mammalian reproduction-specific proteins of the present invention.
 CC
 CC Sequence 3649 BP; 947 A; 841 C; 911 G; 950 T; 0 other:
 CC
 CC SQ

Query Match	16.4%;	Score 381;	DB 22;	Length 3649;
Best Local Similarity	52.7%;	Pred. No. 1.4e-102;		
Matches 975;	Conservative	0;	Mismatches 855;	Indels 21; Gaps 6;

QY	490	CCTGGCTTCACTACTTCCTCCATCTCTCAGTATGAAAAACAGCATCAATGCTCTCGACATGAGTT	549
Db	1	CCTGGCTATCGCGCTAGTATCCGGAGACACAGCGGGGTCTCTCTCTGCTCGTGATGTC	60
QY	550	AGCCATAAAGTCCTTCGAAGTGAGACGTGTTTGGATTTCATGTTCAACITTTATCATCAG	609
Db	61	TCTCATTAAGTGCATTCGGAACGACTCTGTGCTGGATGTCATGCAATCTACTACCAAGCAG	120
QY	610	ACAGAAGAACATAAAATTTCAAGAACAAAGTTTCCAAAGAACTAATAGGTTTAGTTGTTCCT	669
Db	121	AACAAGGAGCAC--TTCCAGGACGAGTGACGACAGCTTCGTGGTGGCAGCATTTGTCATC	177
QY	670	ACCAAGTATTAACAATAAGACATACAGAGTGGATGATATTCAGCTGGGACCCAGAATCCCAAG	729
Db	178	ACGCGCTACACCAATCGTACTACCGAAATCGATGATGTGACTTGGAAACAGACCCCTAAA	237
QY	730	AGCACCTTTAAGAAACCGGACGGCTCTGAAGTCAGCTTCCTTGAATACTACAGAAGCAA	789
Db	238	GACAGCTTTGTCATGTCGGACGGGAAGGAATCACATTCCTCGAATACTACGCAAAAC	297
QY	790	TACAACCAAGAGATCACCGACTTTGAACAGCGCTCTTGTGTCAGCCACAGCCCA--AGAGA	846
Db	298	TATGGGATCACAGTCAAGAGAGATGACCAGCGCTGCTGATCCACGGCCAGTGAGAGA	357
QY	847	AGCGGGGCGCTCGGGGGAACCTGCCAGGCGCTGCCATGCTCAATTCAGAGCTCTGCTAT	906
Db	358	CAGAATAAACCATGGCATGTGTTCTGAAGGGCGAGATCCTGTGCTGCCGAGCTCTCCTTC	417
QY	907	CTTACAGGCTTAAGTAAATTCGTAATGATTTTAACTGATGAAGACITTAGCGGTT	966
Db	418	ATACCGGGGATCCCTGAGAAGATGAAGAAGACTTCAGGGCCATGAAGGACTTGAATCAG	477
QY	967	CATACAAGATAACTCCAGAGCAAAAGCAGCGCTGAAGTGGGAGCAGCTCAATTGATTACATT	1026
Db	478	CAGATTAACTTGAGCCCCAAGCAGCACCGGTGCTTTGGAATGCCCTGCTGCAGAGATT	537
QY	1027	CATAAAAAACGATAATGTTCAAAGGAGCTTCGAGACTGGGGTTTGAGCTTTGATCCCAAC	1086
Db	538	TCAAAAACGAGACAGCCAGCAATGAGCTGACCCGCTGGGGCTCAGTCTGCAATAAGAT	597
QY	1087	TTACTGTCTTCTCAGGAAGAAATTTTGCAACAGAAAGATTCCCAAGGTGGAAAAACA	1146
Db	598	GTCCACAAGATTGAAGGTCGGCTCTTGCCAATGGAGAGGATCAACTTAAGGAACAATTCA	657
QY	1147	TTTGATTACCAATCCAAATTTGAGATTGTTGCCAAGAAACAGAGGTCACCATTAATT	1206
Db	658	TTTGT---CACATCGGAGGCGCTGMACTGGGTTTAAGGAAGTGCACGAGATGCTCCATT	714

Qy 2278 ATTACAGAGAGCCAAATCTGTCTACTGTCAAAACCGCCCTTACTACTCTTAA 2328
Db 1792 TTGCATCATGAGCCAGCATCCAGTGTGTGGAACTGTTCTTCTCTGTAA 1842

RESULT 7

ID ABZ19511 standard; cDNA; 367 BP.
XX AC ABZ19511;

XX 23-JAN-2003 (first entry)

XX Group III cDNA cancer related clone SEQ ID NO:1937.

XX Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
XX immune response; virology; immunology; microbiology; molecular biology;
XX recombinant DNA technology; gene; ss.

XX Homo sapiens.

XX WO200278516-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US10421.

XX 30-MAR-2001; 2001US-280255P.

XX 28-AUG-2001; 2001US-315563P.

XX 09-JAN-2002; 2002US-347313P.

XX (CORI-) CORIXA CORP.

XX Wang T, Wang S, Bangur CS, Gaiger A;

XX WPI; 2003-058387/05.

XX New immunogenic polynucleotides or polypeptides useful for diagnosing,
XX preventing and treating cancer expressing CT or CP mRNA antigens, and
XX in virology, immunology, microbiology, molecular biology and
XX recombinant DNA techniques

XX Claim 1; SEQ ID 1937; 207pp; English.

XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
XX ABP54446 to ABP54472 represent protein (II) sequences, from the present
XX invention. (I) and (II) have cytostatic activity and can be used in gene
XX therapy and vaccines. (I), (II), antibodies and compositions from the
XX present invention are useful for diagnosing, preventing and treating
XX cancer, which expresses CT or CP mRNA antigens. They are useful for
XX stimulating immune response. They can also be useful in virology,
XX immunology, microbiology, molecular biology and recombinant DNA
XX techniques.

XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 367 BP; 118 A; 77 C; 91 G; 81 T; 0 other;

Query Match 15.7%; Score 365.4; DB 25; Length 367;
Best Local Similarity 99.7%; Pred. No. 1.6e-98;
Matches 366; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1291 AAAGTTACACAGCCATGGCGATCAATGAGAAAGCCATTAATGATTGAAGTGATGAC 1350
Db 1 AAAGTTACACAGCCATGGCGATCAATGAGAAAGCCATTAATGATTGAAGTGATGAC 60

Qy 1351 AGAACTGAAGCCTTAAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTT 1410

Db 61 AGAACTGAAGCCTTAAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTT 120

Qy 1411 GTCTGTCTGTGTCAAGTAATCGAAGACAAATACGATGCTATTAAAAAATACCTGTGT 1470

Db 121 GTCTGTCTGTGTCAAGTAATCGAAGACAAATACGATGCTATTAAAAAATACCTGTGT 180
Qy 1471 ACAGATTGCCCTTACCCCAAGTCAAGTGTGTGTGTCGCCCGAACCCTTAGCCAAACAGCAACT 1530
Db 181 ACAGATTGCCCTTATCCCAAGTCAAGTGTGTGTGTCGCCCGAACCCTTAGCCAAACAGCAACT 240
Qy 1531 GTCATGGCCATTGCTCAAAAGATTGCCCTACAGATGAATGCAAGATGGAGGAGAGCTC 1590
Db 241 GTCATGGCCATTGCTCAAAAGATTGCCCTACAGATGAATGCAAGATGGAGGAGAGCTC 300
Qy 1591 TGGAGGTGGACATCCCCCTGAAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGAC 1650
Db 301 TGGAGGTGGACATCCCCCTGAAAGCTCGTGATGATCGTTGGCATCGATTGTTACCATGAC 360
Qy 1651 ATGACAG 1657
Db 361 ATGACAG 367

RESULT 8

ABZ20005
ID ABZ20005 standard; cDNA; 367 BP.
XX AC ABZ20005;

XX 23-JAN-2003 (first entry)

XX Group III cDNA cancer related clone SEQ ID NO:2431.

XX Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
XX immune response; virology; immunology; microbiology; molecular biology;
XX recombinant DNA technology; gene; ss.

XX Homo sapiens.

XX WO200278516-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US10421.

XX 30-MAR-2001; 2001US-280255P.

XX 28-AUG-2001; 2001US-315563P.

XX 09-JAN-2002; 2002US-347313P.

XX (CORI-) CORIXA CORP.

XX Wang T, Wang S, Bangur CS, Gaiger A;

XX WPI; 2003-058387/05.

XX New immunogenic polynucleotides or polypeptides useful for diagnosing,
XX preventing and treating cancer expressing CT or CP mRNA antigens, and
XX in virology, immunology, microbiology, molecular biology and
XX recombinant DNA techniques

XX Claim 1; SEQ ID 2431; 207pp; English.

XX ABQ17575 to ABQ20506 represent isolated polynucleotide (I) sequences, and
XX ABP54446 to ABP54472 represent protein (II) sequences, from the present
XX invention. (I) and (II) have cytostatic activity and can be used in gene
XX therapy and vaccines. (I), (II), antibodies and compositions from the
XX present invention are useful for diagnosing, preventing and treating
XX cancer, which expresses CT or CP mRNA antigens. They are useful for
XX stimulating immune response. They can also be useful in virology,
XX immunology, microbiology, molecular biology and recombinant DNA
XX techniques.

XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 367 BP; 118 A; 78 C; 89 G; 79 T; 3 other;

Query Match 15.6%; Score 364; DB 25; Length 367;
Best Local Similarity 99.2%; Pred. No. 4.1e-98;
Matches 364; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1291 AAAGTTACACAGCCATGGCATGCAAAATGAGAAAGCAATAATGATTGAAGTGGATGAC 1350
DB 1 AAAGTTACACAGCCATGGCATGCAAAATGAGAAAGCAATAATGATTGAAGTGGATGAC 60

QY 1351 AGAAGTGAAGCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATGTT 1410
DB 61 AGAAGTGAAGCTACTTAAGAGTCTTACAGCAAAAGGTCACAGCAGACACCCAGATGTT 120

QY 1411 GTCGTCTGTGTCAGTAATCGGAAGACAAATACGATGCTATTAAAAATACCTGTGT 1470
DB 121 GTCGTCTGTGTCAGTAATCGGAAGACAAATACGATGCTATTAAAAATACCTGTGT 180

QY 1471 ACAGATTCGCCCTACCCCAAGTCAGTGTGTGGCCGAACTTAGGCAACACGCAAACT 1530
DB 181 ACAGATNNCCCTACCCCAAGTCAGTGTGTGGCCGAACTTAGGCAACACGCAAACT 240

QY 1531 GTCATGGCCATGTCACAAAGATTGCCCTCAGATGAACTGCAAGATGGGAGAGGCTC 1590
DB 241 GTCATGGCCATGTCACAAAGATTGCCCTCAGATGAACTGCAAGATGGGAGAGGCTC 300

QY 1591 TGGAGGTGGACATCCCTCGAGCTCGTGATGATCGTTGGCATCGATTGTCATGAC 1650
DB 301 TGGAGGTGGACATCCCTCGAGCTCGTGATGATCGTTGGCATCGATTGTCATGAC 360

QY 1651 ATGACAG 1657
DB 361 ATGACAG 367

RESULT 9
AAH15960
ID AAH15960 standard; cDNA; 2272 BP.
XX
AC AAH15960;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:14558.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 14558; 2537pp + CD ROM; English.
XX

CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX
SQ Sequence 2272 BP; 609 A; 509 C; 585 G; 569 T; 0 other;

Query Match 14.2%; Score 330; DB 22; Length 2272;
Best Local Similarity 51.6%; Pred. No. 1.9e-87;
Matches 884; Conservative 0; Mismatches 810; Indels 18; Gaps 5;

QY 629 AAGAACAGTTTCCAAAGAACTAATAGTGTAGTGTGTTTACCAAGTATAAACAAGA 688
DB 1 AGGATGAGTGTACTAAGCTTCTGGTGGCAATATGTTATCAGCTGATATAACAATCGTA 60

QY 689 CATACAGAGTGGATGATATTGACTGGGACAGCAATCCCAAGAGCAGCCTTTAAAGAAAGCCG 748
DB 61 CCTATGCTATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120

QY 749 ACGGCTCTGAAGTCAGCTTCTTAGAATACCTACAGGAGCAATACAAACAGAGATCACCG 808
DB 121 ATGGGAAAGAGATCACATCTTGGATACCTACAGCAAAATTTATGGATCACAGTTAAGG 180

QY 809 ACTTGAAGAGCCTGTCTTGGTCAGCCAGCCCA---AGAGAAGCGGGGGCTTGGGGGGA 865
DB 181 AAGAGGACCAAGCCATGCTGATTCAGAGCCAGGTCAGAGACAGGATAATCATGGGATGC 240

QY 866 CACTGCCAGGGCCTGCGCATGCTCATCTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 925
DB 241 TGCTAAAGGGGAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300

QY 926 AAATGCGTAATGATTTTAAAGTGAAGAAAGTTCAGCGTTTCATACAGACTAACTCCAG 985
DB 301 AGATGAAGAGGACTTCAGAGCCATGAAGGATTTGGCTCAGCAAAATCAATCTGAGCCCA 360

QY 986 AGCAAGGCGAGCGTGAAGTGGGAGCGACTCATTCGATTACATTCATAAAGAGATATGTC 1045
DB 361 AGCAACACCATAGTGTCTTGGAAATGCTGCTGCAAAAGATTCGAAAGACGAGCGCA 420

QY 1046 AAGGGGAGCTTCGAGACTGGGGTTTCAGCTTTCGATTCGAACTTACTGCTCTTCAGGAA 1105
DB 421 CCAATGAATGATGCTGCTGGGGCTCCGCTGCAAAAGGATGTACATAAGATTGAAGGAC 480

QY 1106 GAATTTTGAACAGAAAGATTCACCAAGGTGGGAAACATTTGATTGATCAATCCCAAT 1165
DB 481 GTGTTCTGCCAATGAAAGAAATTAACCTTA---AAAAATACTTCGTTTATCACATCTCAGG 537

QY 1166 TTGCAGATGGTCCAAAGAAACAGAGGTGCACCATTAATAGTGTAGCCACTAGATA 1225
DB 538 AACTAACTGGGTTAAGGAAGTAACCAAGAGACCCCTTCCATCTTGACTATCCCATGCA 597

QY 1226 ACTGGCTGTGTATCTATAGCGAAGAAATTTATACAGCAGCAATTCATTCATACAAATC 1285
DB 598 TCAGGGCAGCTTTTATCCCAAGAGAGCAATGACCGGCTCGAGAACTGGTCAACATGT 657

1286	QY	TATTTAAAGTTACACCGACCGATGGCGATGCAAAATGAGAAAAGCAATAATGATGAAGT--	1344
658	Db	TGGAGAAGATAGCCGGCCCCATTGGCATGCGTATGAGCCACCGCGCTGGGTGAACATA	717
1344	QY	-GGATGACAGAACTGAAGCCCTACTTTAAGAGTCTTTACAGAAAAGGTCAACAGCAG-	1396
718	Db	AGGATGACCGAATAGAGACTTATGTCAAGAACATTCAATCCACGTTAGGAGCTGAGGGGA	777
1397	QY	ACACCCAGATAGTTGTCTGTCTGTTCAGTAATCGAAGAGCAAAATACGATGCTATTAA	1456
778	Db	AGATACAGATGGTTGTTTGGATCATCATATGGGCCACGCTGATGATCTTATGGGGCCATCA	837
1457	QY	AAAPAACTGTGTGATACAGATTGCCCTACCCCAAGTCAGTGTGTGGTGGCCCGCAACCTTAG	1516
838	Db	AGAACTGTGTGTGTGAGTCCCCAGTGCCTCCCAAGTTGTCAATGTTTGGAAACCAATTG	897
1517	QY	GCAAAACAGAAAACGTCTCATGGCCAATTGCTACAAAGATTGCCCTACAGATGAACCTGCAAGA	1576
898	Db	GTCAAGCCCAACGAGCTTCGGAGTGTGCCCGCAGAAAGATTTACTTCAGATTAACCTGTAAT	957
1577	QY	TGGAGGAGAGAGCTCTCGAGGGTGACATCCCCCTGAAGCTCGTGATGATCGTTGGGATCG	1636
958	Db	TGGGTGTGAGCTCTGGGAGTGGAATATCTCTGAAACAGTTAATGGTGATCGGGATGG	1017
1637	QY	ATTGTTTACCATGATACAGACTGGGGGAGGTCAATCGCAGGATTTGTTGCCAGCATCA	1696
1018	Db	ATGTTTACCATGACCCCGATGAGGATCGCTCCGTGGTTGCTGTCGCAAGCATCA	1077
1697	QY	ATCAAGGATGACCCCGCTGGTTCTCAGCTGTCATATTTTCAGGATGAGAGCAGAGCTGG	1756
1078	Db	ATCTCAACCTTCACAAAATGATATTTCCGGGTGGTGTTCAGATGCCGCATCAGGAGATTG	1137
1757	QY	TAGATGGGCTCAAAGTCTGCTGCAAGCGGCTCTGAGGGCTTGGAATAGCTGCAATGAGT	1816
1138	Db	TGACACGCTGAAGTATGCTCGTGGGCTCTTAAAAAAGTTTATGAGGTGAACCACT	1197
1817	QY	ACATGCCACCGCGATCATCGTGTACCGCATGCGTAGAGAGCGGCCAGCTGAAAACAC	1876
1198	Db	GTCTACAGAGAAGATTGTGTGTACCGTGATGAGTGTCTGATGGCCAACTGAGACAG	1257
1877	QY	TGGTGAACTACGAAGTCCCAAGTTTTTGGATTGTCTTAAATCAATTTGGTAGAGTTACA	1936
1258	Db	TTGCCAACTATGAGATTCCTCAACTACAGAAAGTGTTTGAAGCTTTTGAGA--ATTATC	1314
1937	QY	ACCTGAGCTAAACGGTAATTTGTGTGAGAAAAGAGTGAACACAGATTTTTTGTCTCAGT	1996
1315	Db	AGCCCAAGATGGTGGTGTGTGTAGTCAGAGAAAATCAGTACTAATCTATATCTGGCTG	1374
1997	QY	CTGGAGAGAGACTTCAGAAATCCACTTCCTCGAAACAGTTATTGTATGATAGGTTTACCAAG	2056
1375	Db	CTCCTCAGAACTTTGTAACCTCCCACTCTCTGGAAGCTGTGTAGATCATCAATATACAGCT	1434
2057	QY	CAGAAATGGTATGACTTTTATCGTAGAGCAGGCTGTGAGAGTGGTGTCTTCTCCCA	2116
1435	Db	GTGAGTGGGTGGATTTCTATCTCTTCTGCCCATCATGTACGGCAGGGCTGTGGCATTCCTA	1494
2117	QY	CACATTACATGTCACTATGACAAACAGCGGCTGAGCCAGACCAATACAGCGCTTGA	2176
1495	Db	CGCATATGCTGTGTCTCAACACCGCAACCTGAGCCCTGATCATATGACAGGGGTGA	1554
2177	QY	CCTACAAGCTGTGCCACATCTATTACAACCTGGCCAGGTGTCTTGGTGTCTCTGCTCCTT	2236
1555	Db	CTTTCAAACTGTGCCCATGATCTGGAAATTTGGCTTGGCACTACAGAGTTCACAGCTCCTT	1614
2237	QY	GCCAGTACGCCCAACAGCTGGCTTTTCTGTTGGCCAGAGTATTTCACAGAGAGCCAAATC	2296
1615	Db	GCAAGTATGCCCAACAGCTAGCTTCTGTGACGAGACATCTTCGATCATGACACAGCCA	1674
2297	QY	TGTCACGTCAAAACCGCTTTATACCTCTAA	2328
1675	Db	TCCAGCTGTGCGAGAACTGTGTTCTCTCTGTGA	1706

```

RESULT 10
ABLI0571
ID ID ABLI0571 standard; cDNA; 2838 BP.
XX AC ABLI0571;
XX AC
XX DT 26-MAR-2002 (first entry)
XX XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 26195.
XX DE
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX XX
XX OS Drosophila melanogaster.
XX OS
XX PN WO200171042-A2.
XX PN
XX PD 27-SEP-2001.
XX PD
XX PF 23-MAR-2001; 2001WO-US09231.
XX PF
XX PR 23-MAR-2000; 2000US-191637P.
XX PR
XX PR 11-JUL-2000; 2000US-0614150.
XX PR
XX PA (PEKE ) PE CORP NY.
XX PA
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX PI
XX DR WPI: 2001-656860/75.
XX DR
XX DR P-PSDB; ABB66468.
XX DR
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PT
XX PS Claim 1; SEQ ID NO 26195; 21pp + Sequence Listing; English.
XX PS
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
XX CC sequences (ABLI01840-ABLI16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX CC
XX SQ Sequence 2838 BP; 783 A; 659 C; 682 G; 714 T; 0 other;

Query Match 13.3%; Score 310.6; DB 23; Length 2838;
Best Local Similarity 49.1%; Pred. No. 1.4e-81;
Matches 1134; Conservative 0; Mismatches 1129; Indels 46; Gaps 10

QY 52 AAACAGAGTCTTCAGGCATTATAGTAAGTTAAGCACTAACCATTTCCGGCTGACATCC 111
Db 431 AAGAAGAGGTGGTGGCACTCATATTACCGTCAGGCAACTATTTTAAAGGTATTAAAG 490
QY 112 CTTCCCGAGTGGCGTTATATCAGTATCAATTGACTATTAACCCACTGATGGAAGCGAGA 171
Db 491 CGTCCAAACTGGACCATCTACCACTACCGGTGCGATTTTACGCGCTGATGTGGAGGTACA 550
QY 172 AGACTCCGTTCCAGTCTTCTTTTCAACAGAGATCTTAATTTGGAAGTGTCTATGCTTTT 231
Db 551 CGACATGCGACCGGTCTTCTTGTATGAACATTAAGGGATCTCGGC--GGCTACATCTTGG 608
QY 232 GATGGAACGATATTATTATTTTACCTTAAAGACTACAGCAMA-----AGTTTA 277
Db 609 ACGGAACCAATATGTTTGTGATCAATCAGTTCGAAGCTGTTCAAGATAGCCCTATGTTT 668
QY 278 CTGAAGTTTTTAGTAAAGCCGGAAATCGAGAGGATGTGAGGATTAACGATCACTTTAAACA 337

```

669	TGGA	ACTG	TTAC	GAGAG	AGT	CGT	GCT	TGCG	GAAAA	CATT	GAAT	TTAA	GAT	CA	AGG	CTG	TTG	728	
338	ATGA	ACTT	CCAC	TACAT	CA	CAAA	CTT	GTG	TGC	AGT	TCT	TAAT	ATAT	TAT	TTT	TC	AGG	AGC	397
729	GATC	TG	CAAT	CTTA	CGGAT	CGG	AGAG	TGTT	CAG	GTTC	TAAT	TCT	CA	TACT	TCG	CG	AGG	789	
398	TTTT	GA	AAAT	CAT	GAAT	TTG	CAAC	CAAA	TAT	TGAC	GAAA	TAT	TAT	TA	ACC	CAAT	AGCC	AA	457
789	CCAT	TG	AGG	GGCT	TAG	ACT	TGA	AGT	TGG	TCT	CG	CT	ACT	ACT	AG	CT	CT	CA	848
458	TTGA	TAT	TCCA	AGT	CA	CAG	TTG	TG	TAT	TGG	CC	TG	CTT	CAC	TACT	TC	CA	T	517
849	TTAA	TTT	TGG	AGAA	TTT	CCG	ATG	CAAT	TAT	TG	CC	TG	CT	AT	FC	AG	TTC	GC	908
518	ATGA	AAA	CAG	CAT	CAT	GCT	CTG	CAC	TG	AG	CTT	TAG	CCAT	ATA	AGT	CTT	TG	CA	577
909	ACGA	AAT	GAT	AT	TACT	TTG	TCC	AGAT	TAT	CC	CA	AG	TTT	AT	CG	CA	CT	AG	968
578	TTTT	TG	GAAT	TT	CA	TTT	TAT	CAT	CAG	CAG	AGAA	CA	TAAA	TTT	CA	GA	CA	AG	637
969	TGT	ACA	AT	AT	TTT	TCC	GA	TCT	AT	TCG	CA	GAC	TG	AC	GTAT	CA	AA	GT	1028
638	TTT	CCA	AG	ACT	TA	AT	AG	TTT	AGT	TTT	AGT	TTT	CTT	CA	AG	AT	TA	CA	697
1029	GT	-----	GC	AGT	AT	GG	TAT	GGT	AT	ACT	PA	CT	PA	AC	GG	AT	TA	TA	1082
698	TG	AT	GAT	AT	T	TG	ACT	GG	CA	CC	AG	AG	CA	CTT	TA	GA	AA	AG	757
1083	TTG	AC	GA	TG	CG	ACT	TTT	CA	AT	CG	AC	GC	AT	TG	CA	AA	TTT	---	1139
758	AAGT	CAG	CTT	CTT	AGA	AT	CTA	CTA	AGG	AG	CA	ATA	CA	AA	AG	AG	AT	CAC	817
1140	AGA	TTT	CGT	AC	TG	GA	TTACT	ATA	TAA	GA	AG	CG	ATA	CA	CA	TAA	T	CAT	1199
818	AGC	CTG	TT	TG	TC	CAG	CA	CC	CA	AG	AG	AGG	CGG	CG	CTT	TA	GA	AA	877
1200	AAC	CT	TG	GT	CA	TG	TC	---	GT	CC	AC	AG	CA	AAA	AT	ATT	CGT	GGT	1258
878	CTG	CA	TG	CT	CA	TT	CT	CT	TAC	AG	GT	CT	AA	CT	GA	TA	AA	AT	937
1259	ATA	---	AT	GAT	CA	TT	CCG	AG	CT	GG	CA	CGG	CT	AC	GG	GA	T	AT	1316
938	ATTT	TA	AG	CT	GT	GA	AGA	CTT	AG	CC	GT	TC	AT	CA	CA	AG	CT	AA	997
1317	ACT	TT	AG	CA	CGT	TT	GAG	GG	CA	TAT	AGT	GA	CA	TAC	CA	GG	CT	GA	1376
998	GT	CA	GT	TGG	CA	CG	ACT	CA	T	TA	TT	CA	T	TA	AA	AA	CA	TA	1057
1377	AAG	CTT	CG	AT	TG	TT	CA	CA	AG	CG	CTT	AA	AGT	CGT	TA	AA	CA	AG	1436
1058	GAG	ACT	GGG	GT	TT	GAG	CTT	TG	AT	TCC	AA	CTT	ACT	TG	CT	CT	CT	CA	1117
1437	AGT	CTT	GGA	CA	CT	CA	CG	CT	GA	CT	CG	CT	TG	TG	AG	AT	TC	CA	1496
1118	CAGA	AA	AG	AT	TC	CA	AG	GT	TG	GA	AAAA	CA	ATT	GA	TT	TACA	AT	TC	1177
1497	CGA	AAAA	AT	TAT	TT	TG	GG	CA	CA	AAAA	ATA	TT	CGT	AT	CG	CA	GG	CT	1556
1178	CCAA	AG	AAA	CA	AG	GT	GC	CA	AT	TA	AT	TAG	T	AG	CTT	TA	AG	CA	1237
1557	CCA	AT	G	AG	TT	CG	CA	CT	GT	TT	CG	AT	TT						

RESULT 11
AAZ13404
ID AAZ13404 standard; CDNA; 300 BP.
XX
XX AC
XX AAZ13404;
XX
XX
DT 12-OCT-1999 (first entry)

XX DE Human gene expression product cDNA sequence SEQ ID NO:873.
 XX DE Human; gene; gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX Homo sapiens.
 XX WO9938972-A2.
 XX PD 05-AUG-1999.
 XX PF 28-JAN-1999; 99WO-US01619.
 XX PR 03-APR-1998; 98US-0080666.
 XX PR 28-JAN-1998; 98US-0072910.
 XX PR 24-FEB-1998; 98US-0075954.
 XX PR 31-MAR-1998; 98US-0080114.
 XX PR 03-APR-1998; 98US-0080515.
 XX (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 XX WPI; 1999-494092/41.
 XX Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 XX Claim 1; Page 862-863; 2479pp; English.
 XX The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AA212532 to AA21779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA212532 to AA21779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.
 XX Sequence 300 BP; 75 A; 70 C; 91 G; 64 T; 0 other;
 SQ Query Match 12.3%; Score 286.4; DB 20; Length 300;
 Best Local Similarity 99.3%; Pred. No. 6e-75;
 Matches 298; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 1485 CCCAAGTCAGTGTGTGGTGGCCGACCTTAGGCAACAGCAACTGTCTATGCCATTGCC 1544
 Db 1 CCCAAGTCAGTGTGTGGTGGCCGACCTTAGGCAACAGCAACTGTCTATGCCATTGCC 60
 QY 1545 TACAAAGATTGCCCTACAGATGAACCTAGCAATGGGAGAGAGCTCTGGAGGGGTGGACAT 1604
 Db 61 TACAAAGATTGCCCTACAGATGAACCTAGCAATGGGAGAGAGCTCTGGAGGGGTGGACAT 120

QY 1605 CCCCCTGAAGCTCGTGTGATGATCGTTGGCATCGATTGTTTACCATGATCATCAGCTGGGGC 1664
 Db 121 CCCCCTGAAGCTCGTGTGATGATCGTTGGCATCGATTGTTTACCATGATCATCAGCTGGGGC 180
 QY 1665 GAGGTCAATCGCAGGATTGTTGCCAGCATCAATGAAGGATGACCCGCTGTTCTCAGC 1724
 Db 181 GAGGTCAATCGCAGGATTGTTGCCAGCATCAATGAAGGATGACCCGCTGTTCTCAGC 240
 QY 1725 CTGCATATTTCAGGATAGAGGACAGAGCTGTGTAGTAGGGCTCAAAGTCTGCTGCAAGC 1784
 Db 241 CTGCATATTTCAGGATAGAGGACAGAGCTGTGTAGTAGGGCTCAGAG-CTGCTGCAAGC 299

RESULT 12
 ABL10535
 ID ABL10535 standard; cDNA; 3524 BP.
 XX AC ABL10535;
 XX DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 25087.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 XX PD 27-SEP-2001.
 XX 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR P-PSDB; ABB66432.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Claim 1; SEQ ID NO 26087; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
 CC sequences (ABU1840-ABU16175) and the encoded proteins
 CC (ABU57737-ABU72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 3524 BP; 1053 A; 741 C; 761 G; 969 T; 0 other;
 SQ Query Match 11.3%; Score 263.4; DB 23; Length 3524;
 Best Local Similarity 47.4%; Pred. No. 2.1e-67;
 Matches 1088; Conservative 0; Mismatches 1176; Indels 29; Gaps 9;
 QY 46 GAATCAAAAACAGGTTCTTCAGGCATTATAGAGTTAGCTAACCAATTTCCGGCTG 105
 Db 342 GTATCCCAAGAGGAGCCGATGGCGTCCCGGTCTATGTCGACGACGAACATTTTCCGATTA 401
 QY 106 ACATCCCGTCCCGCTTAT 165

Db 402 AAAACCAAGCCGGAATGGCGGATCGTTTCAATATACGTTGGAGTTTGAGCCGAGCATCGAG 461
Qy 166 GCCAGAAGACTCGGTTCAGTCTTCTTTTCAACAGAAAGTCTAATTCGAAAGTGTAT 225
Db 462 AATCCTCGTGTCCGTATGGAGTTTGTCCATCATCTGCTAACTTCTGGATCAGGCTAT 521
Qy 226 GCTTTTGATGGAACGATATATTTTACCTAAAGACTACAGAAAGGTTTACTGAAGTT 285
Db 522 CTAATTCGATGGAGTCAACTGTTTCCACCAGGAAATTCGAGCGGAATCACTGGTGTCT 581
Qy 286 TTTAGTAAGAACCCGAATGGAGAGGATGAGATAACGATCACTTTTAACAAATGAATCT 345
Db 582 AGCGGAAGTCCAACTGACATTTGAATTAACAAGATATCCATTAAGTTCTGTTGATTCATA 641
Qy 346 CCACCTACATCAACCACTTTGTTGAGTTTCTATAATTAATTTTTCAGGAGGCTTTTGAAA 405
Db 642 TCGTGTGCTAGCCCCGCTTTTTCGAAGTCTTAATCTAATTAATTTGCGCGCTCCATGAAG 701
Qy 406 ATCATGAATTTGCAACAAATTCGACGAATTAATTAACCCAAATGACCCAAATGATTAAT 465
Db 702 GGCCTAAATTTGGAATTAAGTTGGCGGTATCTCTTTGATCCCGAGCTAAGATCGAAATA 761
Qy 466 CCAAGTCACAGGTTGGTGAATTTGGCTGGCTTCACTTCTTCCATCTTTCAGTATGAAGAC 525
Db 762 AGGAGTTCAAAATGAGACTATGGCGGCTATGAGACATCGAATTCGTCAGACGCAAAA 821
Qy 526 AGCATATGCTCTGCACTGACGTTAGCCATAAAGTCTTTCGAAGTGAAGTCTTTTGGAT 585
Db 822 GATATTTTATGGGCAACGAAATACTCACAAAGTTATGGCACCGAGAGCATCTACGAC 881
Qy 586 TTCATGTTCACTTTTATCATCAGACAGAAAGTAAATTTCAAGAACAGTTTCCAAA 645
Db 882 AATAAGCAGGTTGCTCACAAATCCGGCTCTGCTATCAGGACGAAGTACGGGT-----A 935
Qy 646 GAACTAATAGGTTAGTTGTTCTTACCAGTATAAATAAGACATACAGAGTGGATGAT 705
Db 936 AATGTTTGGACTGATGTTCTTACGATTAACAATAACAGAACTTATCGTATCAATGAT 995
Qy 706 ATTGACTGGACCAAAATCCCAAGAGCACTTTTAAAGAACCGAGGCTCTGAAGTCAAG 765
Db 996 GTGACTTTGGACAACTCCGAAATCAACATT--CAGTTGCAAGGTTAGAGATATCAGT 1052
Qy 766 TTCCTAGATATCTACAGGAACATTAACCAAGAGATCAACGACTTGAAGCAGCTGTC 825
Db 1053 TTCGTTGGAATATCTCTCACTAATATATATATACGCAATTCGCGACCAATCAGCCGCTG 1112
Qy 826 TTGTTCAAGCCCAAGAGAGGCGGGCCCTGGGGGACACTCCAGGGGCTCCCAATG 885
Db 1113 CTGATCTCCAAA--AATAGGACAAGGCTCTAATAAATCTAACGCTAGCGAATTAGTGTA 1169
Qy 886 CTCATTCCTGAGCTCTGCTATCTTACAGTCTTAAGTATAAATTCGTAATGATTTAAC 945
Db 1170 CTAATTCCTGAGCTCTGCGAGTGACTGGGCTCAATGCGGAGATCGCTCAAACTTTTCA 1229
Qy 946 GTGATGAAGACTTAGCGTTTACATCAAGACTAATCTCCAGGCAAGGACCGTCAAGTG 1005
Db 1230 CTTATGCGTGCATGAGCAGTTATACGGAATGAACCCCAACACGCACTGATCGATG 1289
Qy 1006 GGAGACTCATTTGATTAATTCATATAAAGCATATGTTTCAAGGGAGGACTTTCGAGACTGG 1065
Db 1290 CGCGCTTTTAAACCCGTTTACAAACACTCCAGAAAGTGTGAGGTTCTTGAGAGACTGG 1349
Qy 1066 GGTTTGAGCTTTGATTCGAATCTGCTTTCTCAGGAAGAAATTTTCAACAGAAAG 1125
Db 1350 AACATGGAACCTGGAACAAGCTCACGAAGTACAAAGCCGGAATTTGGACAGCAGAAC 1409
Qy 1126 AT---TCACCAAGTGGAAACATTTGATTAACAATCCAATTTGAGATTTGTTCCA 1181
Db 1410 ATCGTGTTCATATGGAAGGTTCCCTGCTGGAGAAACCGTGAATTTGGCAAGGCACTTC 1469
Qy 1182 AGAAACAAGAGGTGCAACCAATTAATGATGTTTAAAGCCACTAGATAACTGGCTGTTGATCTA 1241

Db 1470 AGAGACCAAAAGGATGCTTTACCACTCCGAGCGATGGC-CTCGATCGTTGGGCTGTCTATCGC 1528
Qy 1242 TACGCGAAAGAAATATGAAAGCAGCAAAATTCATTCATGATACAAAATCTATTATTAAGTTTACACC 1301
Db 1529 GCCGCAAGGAATTCCTCATGAACTCCGAACTCTACTTGACTCTTTGTATAGAGCAGTAG 1588
Qy 1302 AGCCATGGCCATGCAAAATGAGAAAAGC---AATTAATGATTGAAGTGGATGACAGAACTGA 1358
Db 1589 TGAATGGGTCTTAGAATTCGAAGCCCGCAAGAAATTCATAATTTTATGATGATGACACTGG 1648
Qy 1359 AGCTTACTTAGAGTCTTACAGCAAAAGGTTCAGAGCAGACCCAGATAGTGTCTGCT 1418
Db 1648 AACITATGTGAGAGCAATGAGATGATTTGTGCGCTCAGATCCCAAACTTATATTATGCT 1708
Qy 1419 GTTGTCAAGTAATCGGAAGGACAAATACGATGCTTATTAATAAATAACCTGTCTACAGATTG 1478
Db 1709 CGTACCCNATGATAACGCCGAAGATCTCATCAATCAAAGAAGAGATAGCTTGACAG 1768
Qy 1479 CCTTACCCAAAGTCAGTGTGTGGCCGCAACTTAGGCHAAAAGCAAACTGTCAAGCC 1538
Db 1769 GCGGTGCGCAACTCA---AGTTGTGACCTTAAACCGAACCAAGAACCGTATGATGAG 1825
Qy 1539 CATTCCTACAAAGATTTGCCCTTACAGATCAACTGCAAGTGGAGAGAGCTCTGAGAGGT 1598
Db 1826 CATTCGCCAACCAAAATAGCAATCCAACTGAATGCAAGTTGGGATATACACCTCGATGAT 1885
Qy 1599 GGACATCCCCCTGAAGCTCGTGATGATCGTTGGGATCGATTTGTTTACCATGATGACGACG 1658
Db 1886 CGAACTACCTTGTCCGACTGATGACAAATTTGGCTTTGACATTCGGAAGAGCACACGAGA 1945
Qy 1659 TGGCGGAGGTCAATTCGAGAGATTTGTTGCAAGATCAATGAAGAGGATGACCCGCTGCT 1718
Db 1946 TCGAAGAGGCGCTTACGAGCAATGATTTGCTCAATGATCTACAGCAAAACTCCACGTA 2005
Qy 1719 CTCACGCTGCAATTTTACAGGATAGAGCAGGAGCTGTAGATCGGCTCAAAGTCTGCT 1778
Db 2006 CTTACAGCACAGTACGAGGTGACGCGCTTGTGATGCTCGCTAACACCTTTTGGCCGAT 2065
Qy 1779 GCAAGC---GGCTCTGAGGGCTTGGAAATAGCTGCAATGAGTACATGCCCAAGCCGATCAT 1835
Db 2066 GATAGCAAGGCCCTCGCCAAATCAACATGACATAGGAAGCTGCCATCTCGAATGCT 2125
Qy 1836 CGTGTACCGGATGGCTGAGAGACGCGCACTGAAACACTGCTGTAACACTGACGAAAGTCC 1895
Db 2126 ATTTTATCGAGACGCTGAGCTCGGCTCTCTAAAGCAGCTTTTGAATTTGAAGTCAA 2185
Qy 1896 ACAGTTTTTGGATTTGCTAATAATCCATTGTTAGAGTTACACCTAGACTAAACGTTAAT 1955
Db 2186 GGACATCATTTGAGAGTTGAAACTGAAATAGCCCGGCTCCAGCTAAGCCCAAGCAAT 2245
Qy 1956 TGTGTTGAAGAAAGAGTGAACACAGATTTTGTCTCAGTCTGGAGGAGACTTTCAGAA 2015
Db 2246 AGCT---TATATTGTGTAACTCAATGCAACGCGCTTCTTCTCAACGAGCAAAA 2302
Qy 2016 TCCACTTCTCGAACACTTATGATGTAGAGTTACAGACCAAGATGTTATGATCTTTT 2075
Db 2303 TCCCTCCCTCGTACTATAGTTGATGACGTTATAAAGTCTGCTCCGAGAGATAACGACTTTTA 2362
Qy 2076 TATCGTCAGCCAGGCTGTGAGAGTGTAGTGTCTTCTCCACACATTAACATGTCATCTA 2135
Db 2363 TCTGTTCTCGCAACAGTTCTGTAGGGTACAGTGTGCGGACGAGTACAAATGTTCTTTA 2422
Qy 2136 TGAACAACAGCGGCTGAAGCCAGACCAATACAGCGGCTTGACCTTCAAGCTGTGCCACAT 2195
Db 2423 TAGCAGCATGGTCTCTCACCGGAGAAATGCAAAAACCTTACGTACAGATGTGCCACTT 2482
Qy 2196 CTATTACAACTGGCAGGTGTCTATGTTTCTGCTTCTGCTTGGCAGTACGCCCAAGCT 2255
Db 2483 GTACTCAATTTGTTGCGGCAACCAACAGAGTCCAGAGTGTGCGAGTACGCTAAGAGCT 2542
Qy 2256 GGCCTTTCTGTTGGCCAGAGTATTCACAGAGAGCCAAATCTGTCTACTGTTCACACCGCT 2315
Db 2543 AGCTACCTCTGTTGGTACGAACTTGCACTCTATTCGCAAAAACGCGCTCGAAAAGAGATT 2602


```

Qy 2316 TTACTACCTCTAA 2328
Db 2603 TTAATATCTATA 2615

RESULT 13
ID AAA07586
XX AAA07586 standard; DNA; 3047 BP.
AC AAA07586;
XX 29-AUG-2000 (first entry)
DT XX
DE Drosophila piwi gene.
XX
KW Piwi family protein; piwi; miwi; hiwi; gene therapy; tissue dystrophy;
KW anaemia; immunodeficiency; male infertility; Drosophila; ds.
XX
OS Drosophila sp.
XX
PH Key Location/Qualifiers
FT CDS 84..2615
FT /tag= a
FT /product= piwi
FT /transl_except= (pos:120..122; aa:Xaa)
FT FT /transl_except= (pos:399..401; aa:Xaa)
FT FT /transl_except= (pos:2436..2438; aa:Xaa)
FT FT /note= "Xaa= Leu or Ile"
XX
PN WO200032039-A1.
XX
XX 08-JUN-2000.
XX
XX 03-DEC-1999; 99WO-US28764.
XX
XX 04-DEC-1998; 98US-0110901.
XX
XX (UYDU-) UNIV DUKE.
XX
XX Lin H;
XX
XX WPI; 2000-412085/35.
XX
XX P-PSDB; AAY90233.
XX
XX Piwi family nucleic acids, polypeptides, and antibodies, useful in gene
XX therapy of diseases such as cancer and in various research and
XX diagnostic applications -
XX
XX Claim 19; Page 171-175; 201pp; English.
XX
XX This sequence encodes the Drosophila piwi family protein, designated
XX piwi. The piwi family nucleic acids and polypeptides are used in gene
XX therapy of diseases such as cancer and also in various research and
XX diagnostic applications. The sequences can also be used to treat
XX tissue dystrophy, anaemia, immunodeficiency, and male infertility.
XX
XX Sequence 3047 BP; 936 A; 664 C; 672 G; 772 T; 3 other;

Query Match 11.0%; Score 256.6; DB 21; Length 3047;
Best Local Similarity 47.2%; Pred. No. 2.1e-65;
Matches 1083; Conservative 0; Mismatches 1181; Indels 29; Gaps 9;

Qy 46 GAATCAAAAACAGGTTCTTCAGGCATATATAGTAGGTTAAGCATTACCACTTCCGCGTG 105
Db 342 GTATCCAAAGAGGAAACCGATGGCGTCCCGGTCAATGCTGCAGACGAACTTTTTTCGANTA 401
Qy 106 ACATCCCGTCCCGAGTGGGCTTATATACAGTATCACATTGACTATACCCACTGATCGAA 165
Db 402 AAAACCAAGCGGATGGCGATGTTTCATATACAGTGGAGTTTGGCGACCATCGAG 461
Qy 166 GCGAAGAGATCCGCTTCAGTCTTTCTTTTCAACACGAAAGATCTAAATGGAAAGTGTCA 225

```

```

Db 462 AATCCTCGTGTCCGTATGGGAGTTTTTGTCCAAATCATGCTAACCTTCTGGGATCAGGCTAT 521
Qy 226 GCTTTTGATGGAAACGATATATTTTACCTAAAGACTACAGCAAAAAGTTTACTGAAGTT 285
Db 522 CTATTCGACGGACTGCAACTGTTTCCACCACGAGGAAATTCGAGCAGGAAATCACGGTGTCTC 581
Qy 286 TTTAGTAAGACCCCGAATGGAGAGGATGTGAGATAGCATCTCTTTAAACAAATGAACATT 345
Db 582 AGCGAAAGTTCGAAGCTGGACATTTGAATATCCATAAAGTTTCGTTGGATTCATA 641
Qy 346 CCACCTACATCACCAACTTGTGTGCACTTCTATAATATATTTTTCAGGAGCTTTTGAAA 405
Db 642 TCGTGTGCTGAGCCCGCTTTTTCAGATCTTAATCTTAATTTGCGCGCTCCATGAAG 701
Qy 406 ATCATGAATTTGCAACAAATTTGCAAGAAATTTATTAACCCAAATGACCCCAATGATATT 465
Db 702 GGCTAAATTTGGAATTTAGTTGGCCGTAATCTCTTTGATCCCGAGCTAAGATCGAATA 761
Qy 466 CCAAGTCACAGGTTGGTGATTTGGSCCTGGCTTCACTACTTCCATCTCCATGATGAAC 525
Db 762 AGGAGTTCAAAATGGAGCTATGGCCGGGCTATGAGACATCGATTCGTACGACGNAAAA 821
Qy 526 AGCATCATGCTCTGCACTGAGCTTAGCCATAAAGTTCCTCGAAGTGAGACTGTTTGGAT 585
Db 822 GATATTTTATGGGCACCGAAATAACTCAAAAGTTATGCGCACCGAGACGATCTACGAC 881
Qy 586 TTCAATGTTCAACTTTTATCATCAGACAGAAACATTAATTTCAAGAACAAAGTTTCCAAA 645
Db 882 ATAATGCGAGCTTGTCTCACAAATCGGGCTCGTCATCAGGACGAAAGTACGGGT-----A 935
Qy 646 GAACTAATAGTGTTAGTTGTTCTTACCAAGTATAACAATAAGACATACAGAGTGATGAT 705
Db 936 AATGTTTTCGACTTGAATGTCTTACGATTAATTAACAGAACTTATCGTATCAATGAT 995
Qy 706 ATTGACTGGGACCGAATCCCAAGACACCTTTAAGAAAGCCGACGGCTCTGAAAGTCAGC 765
Db 996 GTCGACTTTGGACAAACTCGAAATCAACATT--CAGTTGCAAGGGTAGAGATATCAGT 1052
Qy 766 TTCTTAGAATACTACAGGAGAAATACACCAAGAGATACCGACTTGAAGACGACCTGTCTC 825
Db 1053 TTCGTGGAATACTATCTCACTAAATATAATATACGCAATTCGCGACCACAATACAGCCGCTG 1112
Qy 826 TTGCTCAGCCAGCCCAAGAAAGCGGGGCGCTGGGGGACACTGCCAGGCGCTGCCATG 885
Db 1113 CTGATTTCCAAA---AATAGGCAAGGCTCTAATAAATACCGTAGCGAATTAAGTGGTA 1169
Qy 886 CTGATTTCTGAGCTCTGCTATCTTACAGGTCTAACTGATAAAATGGTATGATTTTAAAC 945
Db 1170 CTAAATCTGAGCTCTGCGAGTGACTGGGCTCAATGCCGAGATCGCTCAAACTTTTCA 1229
Qy 946 GTGATGAAGACTTAGCCGTTTATACAAAGACTTAATCTCCAGAGCAAAAGGCGAGCTGAAGTG 1005
Db 1230 CTTATCGTGCCTACATGACAGCTTATACCGGAATGAACCCCAACAGGCACTGATCGATTG 1289
Qy 1006 GGACGACTCATTCATTACATTAATAAAGCAATGTTTCAAGGGAGGCTTCGAGACTGG 1065
Db 1290 CGGCTTTTAAACCCGTTTACAAAACATCTCCAGAAAGTGTGAAGGCTTTGAGAGACTGG 1349
Qy 1066 GGTTCGAGCTTTGATTTCCAACTTACTGTCTCTCAGGAAGAATTTTTCAAAACAGAAAAG 1125
Db 1350 AACATGGAATCGACAAGAACGTCACAGAAGTACAGGCGGATTAATGGACAGCAGAAC 1409
Qy 1126 AT----TCACCAAGGTGGAAAAACATTTGATTACAATCCAAATTTGCGAGATTGGTCCAA 1181
Db 1410 ATCGTGTTTTATAAATGAAAAGTTTCTGCTGGGAAAAAGCGTGAATGGCAAAAGGCACTTC 1469
Qy 1182 AGAAACAGAGGTGCACCAATTAATTAAGTGTAAAGCCACTAGATAACTGGCTGTGTACTTA 1241
Db 1470 AGAGACCAAGAGATGTTTACCCTCCGCGGATGGC-CTCGATCGTTGGGCTGTCTATCGC 1528
Qy 1242 TAGCGGAAGAAATTAAGAGCAGCAATTCATTGATACAAAATCTATTATAAGTTACACC 1301
Db 1529 GCGCAAGGAATTCCTATGAACCTCGAACTCTACTTGACTCTTTGTATAGACAGCTAG 1588

```

1170 AGATTGGTCCAAAGAAACAAGAGGTGCACCATTAATTAGTGTAAAGCCACTAGATAACTG 1229

```
Qy 1230 GCTGTTGATCTATAGCGGAGAAATTTATGACGAGCAATTCATTGATACAAATCTATT 1289
Db 63 GGCACCTTTTACCAGAGAGCAATGACGAGCTCGAGAACTGGTCAACATGTTGGA 122
Qy 1290 TAAAGTTACACGAGCATGGGATGCAATGAGAAAGCAATTAATGATTGAAGT---GGA 1346
Db 123 GAAGATAGCGGCCCCATTTGGCATGGTATGAGCCACCGGCTGGTTGAATTAAGGA 182
Qy 1347 TGACAGAACTGAAGCTTAAAGTCTTACAGCAAAAGGTCAACAGAG-----ACAC 1400
Db 183 TGACCAATAGAGACTTATGTGAGAACCAATTCACACGTTAGGAGCTGAGGGAAGAT 242
Qy 1401 CCAGTAGTGTCTGTCTGTGTTGTCAGTAATCGGAAGGCAAAATACGATGCTATTAAAAA 1460
Db 243 ACAGATGTTGTTGTCATCATCATGGGCCACGCTGATGATCTCTATGGGCCATCAATAA 302
Qy 1461 ATACCTGTGTAAGATGCCCTACCCCAAGTCAGTGTGTGTGGCCCAACCTTAGGCAA 1520
Db 303 GCTGTGCTGTGTGCACTCCCAAGTCCCTCCAGGTGTGTAATGTTGCAACCAATTGGTCA 362
Qy 1521 ACAGCAACTGTCATGGCCATTTGCTACAAAGNTTGCCTACAGATGAAGTCAAGATGG 1580
Db 363 GCCCACCAGGCTTCGAGGTGTGGCCCAAGAAATTTTACTTCAGATTAATGTAATGGG 422
Qy 1581 AGGAGAGCTCTGGAGGTGGACATCCCTCTGAAGCTCGTGTGATGATCGTTGGCATCGATTG 1640
Db 423 TGGTGAGCTCTGGGAGTGGATATTCCTCTGAACAGTTAATGGTGATCGGGATGATGT 482
Qy 1641 TTACATGACATGACAGCTGGGCGAGGTCAATCGCAGGATTTGTTGGCAGCATCAATGA 1700
Db 483 TTACCATGACCCAGTAGAGGATCGCTCGTGTGTTGGCTTCGTGGCAAGCATCAATCT 542
Qy 1701 AGGGATGACCGCTGTTCTCAGCTGCATATTTTCCAGNTAGAGACAGGAGCTGTAGA 1760
Db 543 CACCCTCAAAATGTTATTCGGGTGGTGTTCAGATGCCGATCAGGAGATTTGTGGA 602
Qy 1761 TGGGCTCAAAAGTCTGCTGCAAGCGCTCTGAGGCTTGGATAGTCTCAATGAGTACAT 1820
Db 603 CAGCTGAGCTATGCTCTCGTGGCTCTTAAAGTTTATGAGTGAACCACTGTCT 662
Qy 1821 GCCCAGCGGATCATCTGTGTCGCGATGGCGTAGGACCGCCAGCTGAAACACTGGT 1880
Db 663 ACCAGAGAAGATTTGGTGTACCGGTGAGTGTCTGATGCCAACTGAAGACAGATTGC 722
Qy 1881 GAATACGAAGTGCACAGTTTGTGATGTCATAAATCCATTGCTAGAGGTTACAACCC 1940
Db 723 CAATATGAGATCTCTCAACTACAGAAAGTGTGTAAGCTTTTGAGA---ATTATCAGCC 779
Qy 1941 TAGACTAACGGTAATTTGTTGTCAGAAAGAGTGAACACCAAGATTTTGTCTCAGTCGG 2000
Db 780 CAAGATGCTGTTGTTGTTAGTTCAGAGAAATCAGTACTAATCTATCTGCTCAGCA 839
Qy 2001 AGGAGACTTCAGATCCACTCTCTGGAACAGTTATGATGAGGTTTACAGACCGA 2060
Db 840 TCAGAACTTTGTAATCTCCACTCTCTGGAACAGTGTGTAGATCATCAATAAACAAGCTGGA 899
Qy 2061 ATGTATGACTTTTATTCGAGCGAGCGCTGTGAGAGTGTGAGTGTGTTTCTCCACACA 2120
Db 900 GTGGGTGATTTTATCTTCTTGCCCATCATGTACGGCAGGCGCTGTGCAATCTCAGCA 959
Qy 2121 TTACAATGTCATCATCAACACAGCGGCTTGAAGCCAGACACATACAGCGCTTGAACCTA 2180
Db 960 TTATGCTGTGTTCTCAACACGCAAACTGAGCCCTGATCATATGACAGGCTGACTTT 1019
Qy 2181 CAAGTGTGCCACATCTATTAACTGCCAGGTGTCAATGTTGTTCTGCTCTCTGCA 2240
Db 1020 CAAACTGTGCCACATGTAAGTGGATTTGGCTTGGCCACCATCAGAGTTCCAGTCTCTGCA 1079
Qy 2241 GTAGCCCAAGCTGGCTTTTCTGTTGGCCAGAGTATTCAGAGAGGCAATCTGTGC 2300
Db 1080 GTATGCCCAAGTAGCTTTTCTGTTCAGACACATCTTGTGATCATGAAACGACCATCA 1139
```

```
Qy 2301 ACTGTCAAAACCGCCTTTTACTACTCTTAA 2328
Db 1140 GCTGTGGAGAACCTGTTCTTCTCTGTGA 1167

RESULT 15
ABZ19102
ID ABZ19102 standard; cDNA; 209 BP.
XX ABZ19102;
AC ABZ19102;
XX
DT 23-JAN-2003 (first entry)
XX
DE Group III cDNA cancer related clone SEQ ID NO:1528.
XX
KW Human; cancer; tumour; therapy; diagnosis; CT antigen; CP antigen;
immune response; virology; immunology; microbiology; molecular biology;
recombinant DNA technology; gene; ss.
XX
OS Homo sapiens.
XX
FN WO200278516-A2.
XX
PD 10-OCT-2002.
XX
PF 28-MAR-2002; 2002WO-US10421.
XX
PR 30-MAR-2001; 2001US-280255P.
PR 28-AUG-2001; 2001US-315563P.
PR 09-JAN-2002; 2002US-347313P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Wang S, Bangur CS, Gaiger A;
XX
DR WPI; 2003-058387/05.
XX
New immunogenic polynucleotides or polypeptides useful for diagnosing,
preventing and treating cancer expressing CT or CP mRNA antigens, and
in virology, immunology, microbiology, molecular biology and
recombinant DNA techniques
Claim 1; SEQ ID 1528; 207pp; English.
XX
AB017575 to AB020506 represent isolated polynucleotide (I) sequences, and
ABP54446 to ABP54472 represent protein (II) sequences, from the present
invention. (I) and (II) have cytostatic activity and can be used in gene
therapy and vaccines. (I), (II), antibodies and compositions from the
present invention are useful for diagnosing, preventing and treating
cancer, which expresses CT or CP mRNA antigens. They are useful for
stimulating immune response. They can also be useful in virology,
immunology, microbiology, molecular biology and recombinant DNA
techniques.
CC
N.B. The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pot_sequences.
XX
SQ Sequence 209 BP; 48 A; 47 C; 70 G; 44 T; 0 other;

Query Match 8.9%; Score 207.4; DB 25; Length 209;
Best Local Similarity 99.5%; Pred. No. 2e-51;
Matches 208; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1658 CTGGGGGAGGTCAATGCGAGGATTTTTCGCCAGCATCAATGAAGGATGAGCGGTGGT 1717
Db 1 CTGGGGGAGGTCAATGCGAGGATTTTTCGCCAGCATCAATGAAGGATGAGCGGTGGT 60
Qy 1718 TCTCAGCTGCATATTTCAGATAGAGGACAGAGCTGGTAGATGGCTCAAGTCTGGCC 1777
Db 61 TCTCAGCTGCATATTTCAGATAGAGGACAGAGCTGGTAGATGGCTCAAGTCTGGCC 120
Qy 1778 TGCAAGCGGCTCTGAGGCTTGGATAGCTGCAATGATGATGCCAGCGGATCATCG 1837
```

Dd 121 TGCAGGGGCTCTGAGGGCTTGGAATAGCTGCAATGAGTACATGCCCGCCGGATCATCG 180

Qy 1338 TGTACCGCGATGGCGTAGGAGACGGCCAG 1866

Db 181 TGTACCGCGATGGCGTAGGAGACGGCCAG 209

Search completed: December 6, 2003, 12:06:50
Job time : 624 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2003, 11:30:43 ; Search time 4855 Seconds
(without alignments)
11654.134 Million cell updates/sec

Title: US-10-043-774B-1

Perfect score: 2328

Sequence: 1 atgatcttgggtggaacac.....accgccttactactcttaa 2328

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	575.6	24.7	874	10	BF666315
2	515.2	22.1	577	10	BF080918
3	513.4	22.1	527	12	BM751055
4	502.4	21.6	1495	10	BF664941

5	492	21.1	705	10	BF246708	
6	471.2	20.2	533	10	BF080923	
7	459	19.7	3559	11	AK030116 Mus muscu	
8	441.4	19.0	462	9	AA969938 op60ali.s	
9	438.8	18.8	818	12	BG773137 602721862	
10	433.2	18.6	642	13	BU854142 AGENCOURT	
11	415.2	17.8	565	10	BG101107 uyl9e03.y	
c	12	413	17.7	413	9	AL041473 DKEZP434E
13	405.6	17.4	779	12	BI560710 603254744	
14	400.4	17.2	664	13	BU234973 603411470	
15	390.6	16.8	1092	12	BI463482 603203560	
16	388.8	16.7	420	13	BI453482 603203560	
17	387	16.6	587	9	BQ374910 MK4-TN010	
18	373	16.0	935	10	AA430311 zw68h01.r	
19	370.8	15.9	608	13	BU247005 601855750	
20	365.4	15.7	767	13	BU230168 60397851	
21	350.6	15.1	875	14	BU232120 603792231	
22	347	14.9	882	12	CA471292 AGENCOURT	
23	324.6	13.9	715	10	BI459464 603200456	
24	307.8	13.2	666	12	BG721550 602695155	
25	305.8	13.1	622	9	BI463897 603202635	
26	301	12.9	731	12	AL705072 DKEZP686E	
27	296.8	12.7	640	9	BI458664 603199792	
28	296	12.7	833	13	AL705301 DKEZP686G	
29	285.6	12.3	767	14	BU275338 603532885	
30	285.2	12.3	609	13	CB592670 AGENCOURT	
31	275.8	11.8	559	12	BQ481304 faa74e09.	
32	272	11.7	807	13	BM728018 UI-E-EJ0-	
33	265	11.4	617	13	BU160046 AGENCOURT	
c	34	264.8	11.4	512	BJ040023 BJ040023	
35	258.4	11.1	477	12	BM680592 UI-E-EJ0-	
36	249	10.7	718	13	BM539361 h008e02.g	
37	245.2	10.5	568	13	BQ435882 AGENCOURT	
38	243.4	10.5	785	10	BQ074332 f277h06.y	
39	242.2	10.4	378	10	BG221179 RST40982	
40	241.2	10.4	534	12	BF764928 IL2-CS004	
41	241.2	10.4	800	10	BI672860 flt38409.y	
42	240.2	10.3	796	10	BG206124 RST25559	
43	238.2	10.2	778	10	BG209773 RST29299	
44	238.2	10.2	783	10	BG210801 RST30349	
45	238	10.2	795	10	EG198827 RST18097	
					BG191694 RST10790	

ALIGNMENTS

RESULT 1
BF666315
LOCUS 602119233F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4276136 5',
DEFINITION 874 bp mRNA linear EST 21-DEC-2000
mRNA sequence.
ACCESSION BF666315
VERSION BF666315.1 GI:11940210
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 874)
NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCMI096 row: b column: 09
High quality sequence stop: 594.


```

Db      242  GTCTTACAGCAAAAGGTTACATCGGATACCCAGATAGTTGTTGTGTGTCAAGTAAT 301
Qy      1432 CGGAAGGACAAATACATGCTATTAAAAATACCTGTGTACAGATTGCCTACCCCAAGT 1491
Db      302  CGAAGGACAAATATGATGCTATTAAAAATACCTGTGTACAGATTGCCTACTCCAAAT 361
Qy      1492 CAGTGTGTGTGGCCCGCAACCTTAGCAGCAACACGAACTGTGTATGCTGCTTCAAAAG 1551
Db      362  CAGTGTGTGTGGCCCGCAACCTTAGCAGCAACACGAACTGTGTATGCTGCTTCAAAAG 421
Qy      1552 ATTGCCCTACAGATGAATCGAGATGGAGGAGAGCTCTGAGGGTGGACATCCCCCTG 1611
Db      422  ATGCCCTGCGAGATGAATCGAGATGGAGGAGAGCTCTGAGGGTGGACATCCCTGCTTA 481
Qy      1612 AAGCTCGTGATGATGCTTGGCATGATGCTTACCATGATCATGACAGCTGGGGGAGGTCA 1671
Db      482  AAGCTAGCGATGATGCTTGGCATGATGCTTACCATGATCATGACAGCTGGGGGAGGTCA 541
Qy      1672 ATGCCAGGATTTGTTGCCAGCATCAATGAAGGGATG 1707
Db      542  ATTGCAGGATTTGTTGCCAGCATCAATGAAGGGATG 577

RESULT 3
BM751055
LOCUS      K-EST0026956 S9SNU601 Homo sapiens cDNA clone S9SNU601-21-E08 5',
DEFINITION mRNA sequence.
ACCESSION BM751055
VERSION    BM751055.1 GI:19080673
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 527)
AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
           Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
           Kim,Y.S.
TITLE     21C Frontier Korean EST Project 2001
JOURNAL   Unpublished
COMMENT   Contact: Kim YS
           Genome Research Center
           Korea Research Institute of Bioscience & Biotechnology
           52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
           Tel: +82-42-860-4470
           Fax: +82-42-860-4409
           Email: yongsung@mail.kribb.re.kr
           Plate: 21 row: E column: 08
           High quality sequence stop: 527.
           Location/Qualifiers
             1. 527
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="S9SNU601-21-E08"
               /sex="M"
               /tissue_type="Ascites"
               /cell_type="Epithelial"
               /cell_line="SNU-601"
               /lab_host="Top10P"
               /clone_lib="S9SNU601"
               /note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI;
               Site 2: XhoI; The poly (A)+ RNA was dephosphorylated with
               bacterial alkaline phosphatase (BAP) and then dephosphorylated
               with tobacco acid pyrophosphatase (TAP). The decapped
               intact mRNA was ligated with DNA-RNA linker including sf11
               site by treatment of T4 RNA ligase and the first strand
               cDNA was synthesized with Superscript II using Sf11
               oligo-dT primer. After first strand synthesis, RNA was
               degraded by NaOH treatment and cDNA was amplified by PCR
               reaction. The PCR products were digested with Sf11 and
               cloned into DraIII- digested pME18S-FL3 vector. The

```

obtained cDNA vectors were used for transformation of competent cells E. coli Top10P' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

```

BASE COUNT      188 a      95 c      113 g      131 t
ORIGIN
Query Match      22.1%; Score 513.4; DB 12; Length 527;
Best Local Similarity 99.6%; Pred. No. 5.3e-126;
Matches 525; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy      930  GCCTAATCATTTTAAAGCTGATGAAAGACTTAGCCGTTTCATCAAGACTTAATCCAGAGCA 989
Db      1    GCGTAATGATTTTAAAGCTGATGAAAGACTTAGCCGTTTCATCAAGACTTAATCCAGAGCA 60
Qy      990  AAGCAGCGTGAAGTGGGAGCACTCATTGATTACATTCATATAAAGCGATAATGTTCAAAG 1049
Db      61  AAGCAGCGTGAAGTGGGAGCACTCATTGATTACATTCATATAAAGCGATAATGTTCAAAG 120
Qy      1050 GGAGCTTCGAGACTGGGGTTTGGCTTTGAGCTTTGATTCCAACTTACTGCTTCTCAGGAGAAAT 1109
Db      121  GGAGCTTCGAGACTGGGGTTTGGCTTTGATTCCAACTTACTGCTTCTCAGGAGAAAT 180
Qy      1110 TTTGCAAAACAGAAAGATTACCAAGGTGGAAAAACATTTGATTACAATCCACAAATTTGC 1169
Db      181  TTTGCAAAACAGAAAGATTACCAAGGTGGAAAAACATTTGATTACAATCCACAAATTTGC 240
Qy      1170 AGATTGTTCCAAAGAAACAAGAGGTGCACCACTTAATTAAGTTAGTGTAAAGCCACTAGATACTG 1229
Db      241  AGATTGTTCCAAAGAAACAAGAGGTGCACCACTTAATTAAGTTAGTGTAAAGCCACTAGATACTG 300
Qy      1230 GCTGTTGATCTATACGCGAAGAAATATGAGCGAGCCCAATTCATTGATACAAATCTATT 1289
Db      301  GCTGTTGATCTATACGCGAAGAAATATGAGCGAGCCCAATTCATTGATACAAATCTATT 360
Qy      1290 TAAAGTTACACACCCCATGGGCATGCCAAATGAGAAAAAGCAATAATGATTGAAGTGGATGA 1349
Db      361  TAAAGTTACACACCCCATGGGCATGCCAAATGAGAAAAAGCAATAATGATTGAAGTGGATGA 420
Qy      1350 CAGAACTGAAGCTTACTTAAGAGTCTTACAGCAAAAGGTTCACAGACACCCAGATAGT 1409
Db      421  CAGAACTGAAGCTTACTTAAGAGTCTTACAGCAAAAGGTTCACAGACACCCAGATAGT 480
Qy      1410 TGCTGTCTCTGTGTCAGATATTC-CGAAGGACAAATACGATGCTATT 1455
Db      481  TGCTGTCTCTGTGTCAGATATTC-CGAAGGACAAATACGATGCTATT 527

RESULT 4
BM751055
LOCUS      602119191F1 NIH MGC_56 Homo sapiens cDNA clone IMAGE:4276247 5',
DEFINITION mRNA sequence.
ACCESSION BM751055
VERSION    BM751055.1 GI:11938836
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
REFERENCE  1 (bases 1 to 1495)
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL   National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT   Unpublished
           Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-remail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: CLONTECH Laboratories, Inc.
           DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov

```


Plate: LLCM1096 ROW: f COLUMN: 24

High quality sequence stop: 511.

FEATURES

Location/Qualifiers

1..1495

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:4276247"

/tissue_type="primitive neuroectoderm"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NIH MGC 56"

/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:"

SfiI (ggcgctcgcc); Site_2: SfiI (ggcattatggcc);

Double-stranded cDNA was prepared from cell line RNA.

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dt(30)BN-3'

(where B = A, C, G and N = A, C, G, or T). Average

insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA)."

476 a 280 c 536 g 203 t

BASE COUNT

ORIGIN

Query Match 21.6%; Score 502.4; DB 10; Length 1495;

Best Local Similarity 94.7%; Pred. No. 8.4e-123;

Matches 553; Conservative 0; Mismatches 26; Indels 5; Gaps 3;

QY 1333 ATGATTGAAGTGGATGACAGAACTCAAGCTACTTAAGAGTCTTACAGCAAAAGGTCA 1392

DB 8 AGGATTGAAGTGGATGACAGAACTCAAGCTACTTAAGAGTCTTACAGCAAAAGGTCA 67

QY 1393 GCAGACACCCAGATAGTGTCTGTCTGTGTCTCAAGTAATCGAAGGACAAATACGATGCT 1452

DB 68 GCAGACACCCAGATAGTGTCTGTCTGTGTCTCAAGTAATCGAAGGACAAATACGATGCT 127

QY 1453 ATTAAAAAATACCTGTGTACAGATTCCTTACCCCAAGTCAGTGTGTGGTGGCCGAC 1512

DB 128 ATTAAAAAATACCTGTGTACAGATTCCTTACCCCAAGTCAGTGTGTGGTGGCCGAC 187

QY 1513 TTAGCAACACAGCAAACTGTCTATGCCATTTGTCTCAAGAGATTCCTTACAGATGA 1572

DB 188 TTAGCAACACAGCAAACTGTCTATGCCATTTGTCTCAAGAGATTCCTTACAGATGA 247

QY 1573 AAGATGGAGGAGAGCTCTGGAGGTGACATCCCTTCAAGCTCGTGATGATCGTTGGC 1632

DB 248 AAGATGGAGGAGAGCTCTGGAGGTGACATCCCTTCAAGCTCGTGATGATCGTTGGC 307

QY 1633 ATCGATTGTTACCATGACATGACAGCTGGCGGAGTCAATCGAGGATTTGTTGCCAGC 1692

DB 308 ATCGATTGTTACCATGACATGACAGCTGGCGGAGTCAATCGAGGATTTGTTGCCAGC 367

QY 1693 ATCAATGAAGGATGACCCCGTGTCTTCAGCGCTGCATATTTCAAGATAGAGGACAGG 1752

DB 368 ATCAATGAAGGATGACCCCGTGTCTTCAGCGCTGCATATTTCAAGATAGAGGACAGG 427

QY 1753 CTGTAGATGGCTCAAGATCTGCTGCAAGCGCTCTGAGGGCTTGAATAGCTGCAAT 1812

DB 428 CTGTAGATGGCTCAAGATCTGCTGCAAGCGG-TCTGAGGGCTTGAATAGCTGCAAT 486

QY 1813 GAGTACATGCCAGCCGATCATCG-TCTACCGGATGCGCTAGAGACGCCAGCTGAA 1871

DB 487 GAGTACATGCCAGCCGATCATCGTTTACCGGATGCGGTTAGAGAGGACGATGT--- 543

QY 1872 AACACTGTGTGAATACGAAGTCCACAGCTTTTGGATTGCTAA 1915

DB 544 AACCCGTTGACACGAGGGCGGCGAGTTTGGTTCCTAAA 587

RESULT 5

BF246708

LOCUS

705 bp mRNA linear EST 14-NOV-2000

DEFINITION

mRNA sequence.

ACCESSION

BF246708

VERSION

BF246708.1

KEYWORDS

EST

SOURCE

ORGANISM

Homo sapiens (human)

REFERENCE

1. (Bases 1 to 705)

AUTHORS

NIH-MGC http://mgi.nci.nih.gov/

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

UNPUBLISHED

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: CLONTECH Laboratories, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM927 row: m column: 02

High quality sequence stop: 572.

FEATURES

Location/Qualifiers

1..705

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/tissue_type="glioblastoma"

/lab_host="DH10B (T1 phage-resistant)"

/clone_lib="NIH MGC 57"

/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:"

SfiI (ggcgctcgcc); Site_2: SfiI (ggcattatggcc);

Double-stranded cDNA was prepared from cell line RNA.

and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dt(30)BN-3'

(where B = A, C, G and N = A, C, G, or T). Average

insert size 1.55 kb (range 0.9-4.0 kb). 12/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA)."

206 a 145 c 194 g 160 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 621; Conservative

0; Mismatches 30; Indels 11; Gaps 10;

QY 1333 ATGATTGAAGTGGATGACAGAACTCAAGCTACTTAAGAGTCTTACAGCAAAAGGTCA 1392

DB 11 AGGATTGAAGTGGATGACAGAACTCAAGCTACTTAAGAGTCTTACAGCAAAAGGTCA 70

QY 1393 GCAGACACCCAGATAGTGTCTGTCTGTCTCAAGTAATCGAAGGACAAATACGATGCT 1452

DB 71 GCAGACACCCAGATAGTGTCTGTCTGTCTCAAGTAATCGAAGGACAAATACGATGCT 129

QY 1453 ATTAAAAAATACCTGTGTACAGATTCCTTACCCCAAGTCAGTGTGTGGTGGCCGAC 1512

DB 130 ATTAAAAAATACCTGTGTACAGATTCCTTACCCCAAGTCAGTGTGTGGTGGCCGAC 189

QY 1513 TTAGCAACACAGCAAACTGTCTATGCCATTTGTCTCAAGAGATTCCTTACAGATGA 1572

DB 190 TTAGCAACACAGCAAACTGTCTATGCCATTTGTCTCAAGAGATTCCTTACAGATGA 249

QY 1573 AAGATGGAGGAGAGCTCTGGAGGTGGAATCCCTTGAAGCTCGTGATGATCGTTGGC 1632

DB 250 AAGATGGAGGAGAGCTCTGGAGGTGGAATCCCTTGAAGCTCGTGATGATCG-TGGC 308

QY 1633 ATCGATTGTTACCATGACATGACAGCT-TGGCGGAGTCAATCGCAGGA-TTGTGTTGCCA 1690

```

Db 309 ATCGATGTTACCATGACATGACAGTCGGGGCGGAGGTCAATCGCAGGATTTTGTGCCA 368
QY 1691 GCATCAATGAAGGGATGACCCGCTGTTCTCAGCTGCTATATTTTCAGATAGACGACG 1750
Db 369 GCATCAATGAAGGGATGACCCGCT-GTTCTCAGCTGCTATATTTTCAGATAGACGACG 427
QY 1751 AGCTGGTAGATGGGCTCAAAGTCTCCCTGCAAGCGGCTCTGAGGCT-TGGAATAGCT-- 1807
Db 428 AGCTGGTAGATGGGCTCAAAGTCTCCCTGCAAGCGGCTCTGAGGCTGTGGATAGCTTG 487
QY 1808 GCAATGATACATGCGCCGCGGAT-CAATGCTGTCACCGATGCGGTAGGAGACGCGCAG 1866
Db 488 CAATGATGATACATGCGCCGCGGATCCATCGTGTACCGAGATGGGTAGCGACGCGCCAG 547
QY 1867 CTG-AAAACACTGGTGAACACTGCAAGTGCACAGTTTTCGATTCCTTAAATCC-ATTG 1924
Db 548 TTGAAAACCACTGTGAACACTGCAAGTGCACAGTTTTCGATTCCTTAAATCCAAATTG 607
QY 1925 GTAGAGTTTCAACACCTAGCTTAACCGTAAATTTGTGTAAGAAAAGAGTGAACACCAT 1984
Db 608 GTAGAGTTTCAACACCTAGAAAACGTTAAATTTGTGTAAGAAAACAGTGAACACAGATT 667
QY 1985 TT 1986
Db 668 TT 669

RESULT 6
BF080923
LOCUS 533 bp mRNA linear EST 18-OCT-2000
DEFINITION 233690 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF080923
VERSION BF080923.1 GI:10874744
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 533)
AUTHORS Fahrenkrug,S.C., Smith,T.P.L., Feking,B.A., Cho,J., White,J.,
Vallet,J., Wise,T., Rohrer,G.A., Pertea,G., Sultana,R., Quackenbush
,J. and Keefe,J.W.
TITLE Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
JOURNAL Mamm. Genome 13 (8), 475-478 (2002)
MEDLINE 22213789
PUBMED 12226715
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR PRIMERS
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCAGTCAAGG
Plate: 53 row: O column: 21
Seq primer: ATTAGGTGACACTATAG.
FEATURES
Location/Qualifiers
1..533
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 2P1G"
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 185 a 93 c 119 g 136 t

```

ORIGIN

```

Query Match 20.2%; Score 471.2; DB 10; Length 533;
Best Local Similarity 92.9%; Pred. No. 1.1e-114;
Matches 494; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 1132 CAAGGTGGAACCAATTTGATTACCAATCCCAATTTGCGAGATTTGGTCCAAAGAACAGAGA 1191
Db 2 CAAGGTGGAACCAATTTGATTACCAATCCCAATTTGCGAGATTTGGTCCAAAGAACAGAGA 61
QY 1192 GGTGACCAATTAATTTAGTTTAAAGCCACTAGATACTGGCTGTTGATCTATATACGGAAGA 1251
Db 62 GGTGACCAATTAATTTAGTTTAAAGCCACTAGATACTGGCTGTTGATCTATATACGGAAGA 121
QY 1252 AATTATGAGCAGCAATTTCAATTGATCAAAATCTTAAAGTTTACACCGCATGGGC 1311
Db 122 AATTATGAGCAGCAATTTCAATTGATCAAAATCTTAAAGTTTACACCGCATGGGC 181
QY 1312 ATGCAAAATGAGAAAGCAATTAATGATTGAAGTGTGACACAGAACTGAGCCCTACTTAAAGA 1371
Db 182 ATACAAATGAGAAAGCAATTAATGATTGAAGTGTGATAGAACTGGAAGCCTATTAAAGA 241
QY 1372 GTCTTACAGCAAAAGGTCACAGCAGACACCCAGATAGTTGTTCTGTTCTCAAGTAAT 1431
Db 242 GTCTTACAGCAAAAGGTTACATCGATACCCAGATAGTTGTTCTGTTCTCAAGTAAT 301
QY 1432 CGAAGGACAAATACGATGCTTATTAATAATACTCTGTGTACAGATTGCCCTACCCCAAGT 1491
Db 302 CGAAGGACAAATATGATGCTTATTAATAATACTTATGTACAGATTGCCCTACTCCAAGT 361
QY 1492 CAGTGTGTGGTGGCCCAACCTTAGGCAACACAGCAAACTGTTCATGGCCATTGTACAAAG 1551
Db 362 CAGTGTGTGGTGGCCCAACCTTAGGCAACACAGCAAACTGTTCATGGCCATTGTCCGACAAAG 421
QY 1552 ATTGCCCTACAGATGAATCAAGATCGGAGGAGCTCTGGAGGGTGGACATCCCCCTG 1611
Db 422 ATGCCCTGACAGATGAATCAAGATCGGAGGAGCTTTGGAGGGTGTATATGCCCTCTA 481
QY 1612 AAGTCTGTGATGCTGTTGGATCGATGTTTACATCATGACATGACAGCTGGGC 1663
Db 482 AAGTACGATGATGATGTTGGCATTGTTATCATGATACACAGCTGGAC 533

```

RESULT 7

```

AK030116
LOCUS 3559 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched
library, clone:4932443D15 product:piwi like homolog 1
(Drosophila)-like, full insert sequence.
ACCESSION AK030116
VERSION AK030116.1 GI:26326094
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kikunai,T., Tashiro,H., Itoh,M.,

```

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Mishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Ogawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

20530913

11076861

4

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L.M., Staubli, P., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, N., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohtsuki, S. and Hayashizaki, Y.

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Nature 409 (6821), 685-690 (2001)

21085660

11217851

5

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team

TITLE

JOURNAL

REFERENCE

AUTHORS

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 3559)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanaoka, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tegan, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

TITLE

JOURNAL

COMMENT

source

Submitted (16-JUL-2001)

Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

1. 3559

location/Qualifiers

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"
/db_xref="PANTOM,DB:4932443D15"
/db_xref="taxon:10090"
/clone="4932443D15"
/sex="male"
/tissue_type="testis"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
142 - 3057

/note="unnamed protein product; piwi like homolog 1 (Drosophila) - like (MGD) | MGIR:1930036, GB | NM_021308, evidence: BLASTN, 99%, match=3559)
putative"

/codon_start=1
/protein_id="BAC26791.1"
/db_xref="GI:26326098"

/translation="MDPVRPLFRGTPVHPSCQVRMPGCPQAPRPLBPAMGRAGPAG RGLVFRPESSPLPQVQKDSVGLVSMFRGMLDTAFRPFKREVPVLLGRVGLRGL SANVRKDRREPSRSLPDSVLAAGDSKLAASVGSVMGRGSEVSLPLGRGAASS IGRWDKPPSAFGLTARDPPRLPQPPALSPSLHSADPPVLTWBERKEILLVKGSK GTPOSGLNLTKI OCHNEAVYQVHTVPSVECKSMRGLKDHQSVGTGVTAFDGS L LPLVKQVVELSKQKTDAAESIKIQLTKILEPCSDLCIPFNVVFRVRKLLDM KLNGRNFYDPSAMVLOQRLLQIWPFGAASIRRTDGLLADVSHKVRMSDLDVM HATIQNKBFHFDKESKLVGSIIVITRYNNRTYRIDVDNKTPKDSFMSDGEKITF LEYKNYGTIVKEDDQPLIHRPESRNNHGMILLKGLILLPELSFTMTGPERMKD FRAMKDTQINLSPKOHGALCCLQRI SONEAASNELTRWGLSLHDKVHKIEGRLL PMERINLRNFTSVTSEDLNVKVEVTRDASILTIPMFHFWALFVPRKAMDQARELVNMLE KIAGIPGRISPPALPKDRIRIETRTQSLGVEGKIQNVCIINGTRDDLLGAI KKLCCVSPSPQVNVRTIQPTLRSAVKILLQNMCKLGBELNGVDIFLQMLVI GMDVHPEKIVVDRSDGQKLVANYEIPLOKQFEAFDNVHPKVVVQVKIST NLILAAPDFHFTVPGTVDHTITSCWVDVFLLAHVROCGGPIPTHYICVCLNTANLS PDHMRUTFKLCHMYNNWPGTIRVPAECKVAHKLAFLSGLLHHEPRLQCLNFFL"

BASE COUNT 908 a 891 c 917 g 843 t

ORIGIN

Query Match 19.7%; Score 459; DB 11; Length 3559;

Best Local Similarity 52.4%; Pred. No. 5.8e-111;

Matches 1161; Conservative 0; Mismatches 1035; Indels 21; Gaps 6;

QY 124 GCCTTATATCATGATCATCTGACTATAACCCACTGATGAAGCAGCAGAACTCCGTTCA 183

Db 850 GCAGTTTATCAGTACCATGTGACTTTTCAGCCCCAGTGTGAATCAAAAGCATGAGGTTT 909

QY 184 GCTCTCTTTTCAACACGAGAGTCTTAATTGGAAGTGTCTGCTTTTGTGGACCATTA 243

Db 910 GGCATGTTGAGGACCACTGCTGCTGCTGGAACCTCACTGCTTTTGTGCTCTATT 969

QY 244 TTAATTTTACCTAAAGACTACAGCAAAAGGTACTGAAGTTTCTAGTAAAGCCCGGAAT 303

Db 970 CTTTATCTCTGTTAAGCTTCAACAGTTGTTGAGTTTAAAGTCAAGAGAACTGAC 1029

QY 304 GGAGAGATGTGAGGATAAGATCACTTAAACAAAGTAACTTCCACCTATCATCACTCAACT 363

Db 1030 GATGCCGAGATCAGTATCAAGATTTCAGTCAACAAAGATCCTGGAGCGGTCTTCGACTG 1089

QY 364 TGTGTCAGTCTTATATATATTTTTCAGGAGGCTTTTGAATCATGAATTTGCACAA 423

Db 1090 TGCATCCCTTCTACAATGTTCTTCCGCGGGGTAAAGAACTTCGAGATATGAAGCTT 1149

QY 424 ATTGGAGAAATTTATATAACCAAAATGACCCAAATGATATCCAACTCAAGTTCAGTTCG 483

Db 1150 GTGGGAGAACTTCTATGACCTCAAGTGCCATGCTACTGCGCAACACAGATTGCAG 1209

QY 484 ATTGGCTGCTCTACTACTTCCATCTTCAATGATGAAGACAGCATCATGCTGCTGCACT 543

Db 1210 ATCTGGCTGGCTATCGGCTAGTATCCGAGGACAGACGGGGTCTCTTCTCTGCTCGCT 1269

QY 544 GAGCTTAGCCATAAAGTCCCTTGAAGTGAGACTGTTTGGATTTCATGTTCACTTTAT 603

Db 1270 GATGTTCTCTAAGGTCAITTCGGAACGACTCTGTGCTGATGATCATGCTATCTATC 1329

QY 604 CATCAGACAGAGACATAAATTTCAAGAACAGATTTTCCAAAGAACTAATAGCTTTAGTT 663

1330 CAGCAGAAAGAGC---ACTTCCAGGACGATGTCAGCAAGCTCTGGTTGGCAGCAATT 1386
 664 GTCTTTACCAAGATATACATAGACATACAGAGTGGATGATATTGACTTGGGACACAGAA 723
 1387 GTATCAGCGGCTACCAACATCTGATCCGATCCGATGATGATGATGATGATGATGATGAT 1446
 724 CCCAAGAGCACCCTTTAAGAAAGCCGCGCTCTGAAGTCTGAGTCTTCTTGAATATCTACAGG 783
 1447 CCTAAAGACAGCTTTGTCTGTCGGACGGGAAGAAATCACATTTCTGGAATACTACAGC 1506
 784 AAGCAATACCAAGAGATACCGACTTGAAGCAGCTGTCTTGGTCAGCCAGCCCA-- 841
 1507 AAAAACTATGGGATCAAGTCAGGAGGAGATGACAGCCGCTGCTGATCCACCGGCCAGT 1566
 842 -AGAGAAGGGGGGCTTGGGGGACACTGCCAGGGCCCTGCCATGCTCAATCTCTGAGTCTC 900
 1567 GAGAGACAGAAATACCATGCGATGTTGCTGAAGGGGAGATCTGCTGCTGCCGAGCTC 1626
 901 TGTATCTTACAGCTCTAATCTGATTAATATGCGTGAATGATTTTAACTGATGAAAGACTTA 960
 1627 TCCTTCATGACGGGGATCCCTGAGAGATGAGAGGAGACTTCAGGGCCCATGAAGGACTTG 1686
 961 GCCGTTTCATACAGACTAATCTCCAGAGCAAGGACGCTGAAGTGGAGACTCTTATGAT 1020
 1687 ACTCAGAGATTAACCTGAGCCCCAAGCAGCACCAGGTCGTTTGAATGCTGCTGAG 1746
 1021 TACATTCATAAAACAGATAATGTTCAAGAGGAGCTTCGAGACTGGGTTTGAAGTTGAT 1080
 1747 AGAATTTCAAAAACAGGAGCAGCAGCAATGAGCTGACCGCTGGGGGGCTCACTGTGAT 1806
 1081 TCCAACTTACTGCTCTCTCAGGAGAAATTTTCCAAACAGAAAGATTCACCAAGTGA 1140
 1807 AAGATGTCACAGATGTAAGTGGCTTCTCCCAATGAGAGGATCACTTAAGAAAC 1866
 1141 AAAACATTTGATTAATTCACAAATTTGCAAGATTTGGTCCAAAGAAACAGAGGTGCACCA 1200
 1867 ACTTCATTG---TCATCCGAGGAGCTGAACTGGTTAAGGAAGTGACAGAGATGCT 1923
 1201 TTAATAGTGTAAAGCACTAGATTAAGTCTGCTGATCTATGAGGAGAAATATGAA 1260
 1924 TCCATTTCAACTATTCCTCATGCTATTTCTGGGCTCTTTTATCCAAAGAGCAATGGAC 1983
 1261 GCAGCAATTCATTGATACAAATCTATTTAAAGTTACACAGCCATGGGCAATGCAATG 1320
 1984 CAAGCCAGAGAACTGGTTACATGTTGGAAGAAAGTTGCCGGGCCATGGCATGGCATA 2043
 1321 AGAAAGCAATAATGATGTAAGT---GGATGACAGAACTGAGCCCTACTTAAGATCTTA 1377
 2044 AGCCCCCAGCCTGGGTGAGCTGAAGGATGACCGAATAGAGACCTATATCAGGACCAAT 2103
 1378 CAG-----CAAAAGGTACAGCAGACACCCAGATAGTCTGCTGTTGTCAAGTAAT 1431
 2104 CAGTCTTACTGGGAGTTGAGGAGAGATACAAATGTCGTTGTCATCATGAGGACCA 2163
 1432 CGGAAGGACAAATACGATGCTATTTAAATAATACCTGTGTAAGATGCTCCCTACCCCAAGT 1491
 2164 CGTGATGATCTCTATGAGGACCATCAAGAGCTGTGCTGCTGAGTCCCGAGTCCCTCA 2223
 1492 CAGTGTGTGGTGGCCGAACTTAGGCAACACAGCAACTGCTCATGGCCATGTCACAAAG 1551
 2224 CAGGTGATCAATGTCCGAACCAATTTGGTCAAGCCCAAGGCTTCGAGCGTGGCTCAGAAA 2283
 1552 ATTGCCCTACAGATGAATCTGACATGGGAGGAGAGCTCTGGAGGTGGACATCCCCCTG 1611
 2284 ATTTTACTTCAGATGAATCTGTAATCGGTGGTGGAGCTCTGGGAGTGGATATCCGCTG 2343
 1612 AAGCTCGTATGATGTTGGCATCGATTTTACCATGACATGACAGCTGGGCGGAGTCA 1671
 2344 AAACAACCTAATGGTGAATGGATGATGTTGATGATGATGATGATGATGATGATGATGAT 2403
 1672 ATCGAGGATTTGTTGCGAGCATCAATGAAGGATGACCGCTGCTTCTCAGCGTGCATA 1731

Db 2404 GTGTCGGCTTCTGGCCAGCATAAATCTCACTACCAATGTTACTCGAGGTGGT 2463
 Qy 1732 TTTTCAGATAGACAGCAGGAGCTGGTAGTGGGCTCAAAAGTCTGCTGCAAGCGGTCTG 1791
 Db 2464 TTTCCAGATGCCACATCAGGAGATTTGGACAGCCTGAGCTGTGCTGGTGGTCTCTTG 2523
 Qy 1792 AGGCTTGGATAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1851
 Db 2524 AAAAAATTTATGAGGTGAACCAATTTGTTCCCAAGAAAAATTTGTTGTTACCGAGATGA 2583
 Qy 1852 GTAGGAGAGCGGCGAGCTGGAACACACTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1911
 Db 2584 GTGCTGATGGCCAGCTTAAGACAGTTGCCAATCTACAGATCCCTCAGCTGCGAAGTGT 2643
 Qy 1912 CTAAATTCATTTGGTAGAGTTTAAACCCCTAGACTAACCGTGAATTTGTTGTTGAAGAA 1971
 Db 2644 TTTGAGCGCTTTGATTAAC---TACCACCCCAAGATGGTGGTGTGTTGTTAGTTTCAAGAAA 2700
 Qy 1972 GTGAACACAGATTTTGTCTGCTGCTGAGTGGAGGAAGACTTCAGAAATCCACTTCTCGAACA 2031
 Db 2701 ATCAGCAACCAATCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2760
 Qy 2032 GTTATTCATGTAGAGGTTTACCAGACAGATGATGATGATGATGATGATGATGATGATGAT 2091
 Db 2761 GTGTTGATCATACCAATACCACTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2820
 Qy 2092 GTGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2151
 Db 2821 GTGCGACAGGCTGTGGCATACCTACCACTGATCTGATCTGATCTGATCTGATCTGATCTG 2880
 Qy 2152 AAGCCAGACACATACAGCGCTGACCTACCAAGCTGTGCCACATCTATTAACAATGGCCCA 2211
 Db 2881 AGCCCTGATCAGATGACAGAGTTGACTTTCAAACTATGCCACATGACTGGAATGGCCT 2940
 Qy 2212 GGTGCTATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2271
 Db 2941 GGTACCATCCAGTTCAGCTTCCAGTATGCTGCAAGTATGCCCAAGCTAGTGGTGGTGG 3000
 Qy 2272 CAGAGTATTCAGAGAGCCAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2328
 Db 3001 CAGATTTGCTATGAGCCAGCCATCCAGTGTGGGAACTGTTCTTCTCTGTAA 3057

RESULT 8
 AA969938
 LOCUS
 DEFINITION
 IMAGE:1581212 3' similar to WP:0105.2 CE06748, mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 521 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham.
 Location/Qualifiers
 1. 462
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1581212"
 /lab_host="DH10B"

/clone lib="Soares NFL T CRC S1"
 /note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
 Equal amounts of plasmid DNA from three normalized
 libraries (fetal lung NBH19W, testis NHT, and B-cell
 NCI CGAP GC81) were mixed, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR-amplified cDNAs from pools of 5,000 clones made
 from the same 3 libraries. The pools consisted of
 I.M.A.G.E. clones 297480-302087, 682632-687239,
 726408-728711, and 729096-731399. Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 Soares 127 a 97 c 124 g 114 t

Query Match 19.0%; Score 441.4; DB 9; Length 462;
 Best Local Similarity 99.6%; Pred. No. 9.4e-107;
 Matches 453; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
 QY 1714 TGGTCTCAGCTGCTATTTTCAGGATAGAGCAGGAGCTGTAGATGGCTCAAAGTC 1773
 Db 1 TGGTCTCAGCTGCTATTTTCAGGATAGAGCAGGAGCTGTAGATGGCTCAAAGTC 60
 QY 1774 TGCCTGCAAGCGGCTCTGAGGGCTTGCATAGCTGCAATGAGTACATGCCCGCCGATC 1833
 Db 61 TGCCTGCAAGCGGCTCTGAGGGCTTGCATAGCTGCAATGAGTACATGCCCGCCGATC 120
 QY 1834 ATCGTGTAACGGATGGCTAGGAGCGGCCAGCTGTAACACTACGAAGTG 1893
 Db 121 ATCGTGTAACGGATGG-GTAGGAGCGGCCAGCTGTAACACTGTAACACGAGTG 179
 QY 1894 CCACAGTTTTTCGATGTCCTAAATCCATTGTAGAGGTACACCCCTAGACTAACGGTA 1953
 Db 180 CCACAGTTTTTCGATGTCCTAAATCCATTGTAGAGGTACACCCCTAGACTAACGGTA 239
 QY 1954 ATTTGGTGAAGAAAGAGTGAACACCCAGATTTTTTGTCTGCTGGGAGGAGCTTCAG 2013
 Db 240 ATTTGGTGAAGAAAGAGTGAACACCCAGATTTTTTGTCTGCTGGGAGGAGCTTCAG 299
 QY 2014 AATCCATTCCTGGAACTTATGATGTAGAGTTTACCAGACCAAGATGGTATGACTTT 2073
 Db 300 AATCCATTCCTGGAACTTATGATGTAGAGTTTACCAGACCAAGATGGTATGACTTT 359
 QY 2074 TTTATCGTGAGCGAGCTGTGAGAGTGTAGTGTTCCTCCACACATTAATGCTATC 2133
 Db 360 TTTATCGTGAGCGAGCTGTGAGAGTGTAGTGTTCCTCCACACATTAATGCTATC 419
 QY 2134 TATGACACAGCGGCTGGAAGCCAGCACATACA 2168
 Db 420 TATGACACAGCGGCTGGAAGCCAGCACATACA 454

RESULT 9
 BG773137
 LOCUS
 DEFINITION 602721862F1 NIH_MGC_97 Homo sapiens cdna clone IMAGE:4838918 5',
 mRNA sequence.
 ACCESSION BG773137
 VERSION BG773137.1 GI:14083790
 KEYWORDS EST.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgsbbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki

Toshiyuki and Piero Carninci (RIKEN)
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10773 row: k column: 15
 High quality sequence stop: 789.
 Location/Qualifiers
 1. 818
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4838918"
 /lab_host="DH10B"
 /clone_lib="NIH MGC 97"
 /note="Organ: testis; Vector: pBluescriptR (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dr primed using primer 5'-TTTTTTTTTTTTTTVN-3',
 size-selected for average insert size 2.2 kb and
 normalized to 500. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carninci, in preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 245 a 174 c 198 g 201 t

FEATURES

Query Match 18.8%; Score 438.8; DB 12; Length 818;
 Best Local Similarity 97.3%; Pred. No. 6.4e-106;
 Matches 500; Conservative 0; Mismatches 7; Indels 7; Gaps 5;
 QY 8 TTGTGTGAACACAGGAGCAACTTAGACCACTGTTAAAGATCAAAAACAGTCTTCAG 67
 Db 305 TTGTGTGAATACAGGAGCAACCTAGACCACTGTTAAAGATCAAAAACAGTCTTCAG 364
 QY 68 GCATTATAGTAAAGTTAAGCACTAACCAATTTCCGGCTGACATCCCGTCCCGAGTGGGCT 127
 Db 365 GCATTATAGTAAAGTTAAGCACTAACCAATTTCCGGCTGACATCCCGTCCCGAGTGGGCT 424
 QY 128 TATATCATGTATCATTTGACTATTAACCACTGATGGAAGCCAGAGATCCGTTCACTC 187
 Db 425 TATATCATGTATCATTTGACTATTAACCACTGATGGAAGCCAGAGATCCGTTCACTC 484
 QY 188 TTCTTTTCAACACGAAGATCTAATTTGGAAGTGTCTATGCTTTTGTGGAACGATATAT 247
 Db 485 TTCTTTTCAACACGAAGATCTAATTTGGAAGTGTCTATGCTTTTGTGGAACGATATAT 544
 QY 248 TTTTACCTAAAGACTACAGAAAGGTTACTGAAGTTTTTTAGTAAGCCCGGAATGGAG 307
 Db 545 TTTTACCTAAAGACTACAGAAAGGTTACTGAAGTTTTTTAGTAAGCCCGGAATGGAG 604
 QY 308 AGGATGTGAGGATAACGATCACTTTTACAAA-TGAACCTTCCACCTACATCAACCACTTGT 366
 Db 605 AGGATGTGAGGATAACGATCACTTTTACAAA-TGAACCTTCCACCTACATCAACCACTTGT 664
 QY 367 TTGCAGTCTTATATATTTTTCAGGAGGCTTTTGAAGATCATGAATTTGTC--AACA 422
 Db 665 TTGCAGTCTTATATATTTTTCAGGAGGCTTTTGAAGATCATGAATTTGTC--AACA 724
 QY 423 AATTGGACGAATTTATTAACCCAAATGACCAA-TTGATATTTCCAGTCAAGTGG 481
 Db 725 AATTGGACGAATTTATTAACCCAAATGACCAA-TTGATATTTCCAGTCAAGTGG 784
 QY 482 TGATTGG-CCTGGCTCACTACTTCCATCCTTC 514
 Db 785 TGATTGGCCCTGGCTCACTACTTTCATCTTCATCCTTC 818

RESULT 10
 BU854142
 LOCUS
 DEFINITION BU854142 NIH_MGC_82 Homo sapiens cdna clone IMAGE:6621314

Db 363 AATGAGTACATGCCAGCGGATCATCGTGTACCGGATGGCTAAGAAAGCGCAGCTG 422

QY 1870 AAAACACTGGT-GAACTACGAAGT-GCCACAGTCTTTTGGATTGTCTAA 1917

Db 423 AAAACACTGGTGGAACTAGCAAGTGGCCACAGCTTTTGGATTGCTAAA 472

RESULT 11

EG1011107

LOCUS

DEFINITION

uvy9e03.y1 McCarrey Eddy spermatoocytes Mus musculus cDNA clone

IMAGE:3658468 5' similar to TR:095404 095404 HIWI ;, mRNA sequence.

ACCESSION

EG1011107

VERSION

EG1011107.1 GI:12596424

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus

REFERENCE

1 (bases 1 to 565)

Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

TITLE

The WashU-NCI Mouse EST Project 1999

JOURNAL

Unpublished

COMMENT

Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LNL ; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

MG1:1420772

Seq primer: Primer name ambiguous

High quality sequence stop: 418.

FEATURES

source

1..565

/organism="Mus musculus"

/mol_type="mRNA"

/strain="CD-1"

/db_xref="taxon:10090"

/clone="IMAGS:3658468"

/sex="male"

/tissue_type="spermatoocytes, pooled from multiple mice"

/dev_stage="60 day"

/lab_host="DH10B (phage-resistant)"

/clone_lib="McCarrey Eddy spermatoocytes"

/note="Organ: testis; Vector: pBluescript SK+ (Stratagene

[5]- Site 1: XhoI; Site 2: EcoRI; cDNA oligo dt-primed

[5']-(GA)10-ACGATCTCGAGTCTTTTCTTTT-3' and directionally

cloned using 5' linkers 5'-AATTCGGCAGCAG-3' and

5'-CTCTGCGC-3'. Size selection of >400bp material gives

average insert size ranging from 1-2 kb. Library was mass

excised (from lambda-UniZAP-XR) and resulting

single-stranded phagemids were prepped and transformed

into DH10B. Library contains 98% recombinants.

References: J. Androl. 20:635-639 and Gene 25:263-269.

Library constructed and donated by J. McCarrey, Ph.D.

(Southwest Foundation for Biomedical Research, Dept. of

Genetics); excision done by E.M. Eddy, Ph.D. (National

Institutes of Health, National Institute of Environmental

Health Sciences). Original lambda-based library is

available through ATCC, catalog #63422."

BASE COUNT 128 a 140 c 168 g 129 t

ORIGIN

Query Match 17.8%; Score 415.2; DB 10; Length 565;

Best Local Similarity 83.5%; Pred. No. 1.1e-99;

Matches 471; Conservative 0; Mismatches 93; Indels Gaps 0


```

QY 1677 AGGATTGTTGCCAGCATCAATGAAGGATGACCCCGCTGTTCTCAAGCTGCATATTTCA 1736
Db 1 AGGATTCGTCGCCAGCATCAATGAAGGATGACCCCGCTGTTCTCAAGCTGCATATTTCA 60
QY 1737 GGATAGAGGACAGGAGCTGTTAGATGGCTCAAGGTCCTGCGAGCGGCTCGAGGC 1796
Db 61 GGACCGGGCGAGGAGCTGTTAGATGGCTCAAGGTCCTGCGAGCGGCTCGAGGC 120
QY 1797 TTGGAATAGCTGCAATGAGTATGATGCCAGCGGATCATCGTGATACCGGATGGGTAGG 1856
Db 121 TTGGAATAGCTGCAATGAGTATGATGCCAGCGGATCATCGTGATACCGGATGGGTAGG 180
QY 1857 AGACGCCAGCTGCAAACTGTTGATGAGTATGAGTATGAGTATGAGTATGAGTATGAGT 1916
Db 181 GGACGGGCGAGCTGAGACCGCTGTTCAATGATGAGTATGAGTATGAGTATGAGTATGAGT 240
QY 1917 ATCCATTGTTAGAGGTTCAACCCCTAGACTTAAGGTTAATGTTGTTGTTGTTGTTGTTG 1976
Db 241 GTCACTCGGAGAGGTTCAACCCAGACTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTG 300
QY 1977 CACCGAATTTTGTCTAGCTGAGGAGAGCTTCAGATCCACTTCCTGGACAGTTAT 2036
Db 301 TGCCAGGTTTTTGTCTAGCTGAGGAGAGCTTCAGATCCACTTCCTGGACAGTTAT 360
QY 2037 TGATCTAGAGGTTACAGACAGATGTTGATGATGATGATGATGATGATGATGATGATGAT 2096
Db 361 CGATGTGAAGTTCACAGACAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 2097 AAGTGTGATGTTCTCCACACATTACAATGATGATGATGATGATGATGATGATGATGATGAT 2156
Db 421 AAGCGGAGTGTGTTCCCAACACATCAATGATGATGATGATGATGATGATGATGATGATGATG 480
QY 2157 AGACACATACAGCGCTTGACCTACAGCTGTGCCACATCTATTACACTGGCCAGGTGT 2216
Db 481 CGACCATTCAGCGCTGACATACAGCTGTGCCACATCTATTACACTGGCCAGGTGT 540
QY 2217 CATTCGTGTTCTGCTCCCTGGCA 2240
Db 541 CATCGAGTTCTGACCTGGCA 564

```

```

RESULT 12
LOCUS AL041473 413 bp mRNA linear EST 29-FEB-2000
DEFINITION DKFZp34E2417 s1 434 (synonym: hres3) Homo sapiens cDNA clone
VERSION AL041473
KEYWORDS AL041473.1 GI:5420824
SOURCE EST.
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Bloecker, H., Boecker, M., Brandt, P., Mewes, H.W., Gassenhuber, J., and
Wiemann, S.
TITLE EST (Bloecker, et al.)
JOURNAL Unpublished
COMMENT Contact: Bloecker H
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 3' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by GBF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
ri sequence also available.
This clone (DKFZp34E2417) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES Location/Qualifiers

```

source

```

1..413
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp34E2417"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="434 (synonym: hres3)"
/note=vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT 100 a 118 c 83 g 112 t
ORIGIN
Query Match 17.7%; Score 413; DB 9; Length 413;
Best Local Similarity 100.0%; Pred. No. 3.6e-99;
Matches 413; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1637 ATTGTTACCATGACATGACAGCTGGGGAGGTCAATCGCAGGATTTGTTGCCAGCATCA 1696
Db 413 ATTGTTACCATGACATGACAGCTGGGGAGGTCAATCGCAGGATTTGTTGCCAGCATCA 354
QY 1697 ATGAAGGATGACCGGCTGTTCTCAGCTGCATATTTTCAGATAGAGGACAGAGCTGG 1756
Db 353 ATGAAGGATGACCGGCTGTTCTCAGCTGCATATTTTCAGATAGAGGACAGAGCTGG 294
QY 1757 TAGATGGGCTCAAAAGTCTGCTGCAAGCGGCTCTGAGGGCTTGGAAATAGCTGCAATGAGT 1816
Db 293 TAGATGGGCTCAAAAGTCTGCTGCAAGCGGCTCTGAGGGCTTGGAAATAGCTGCAATGAGT 234
QY 1817 ACATGCCAGCGGATCATCGTGTACCGGATGCGTAGGAGACGGCCAGCTGAAAAACAC 1876
Db 233 ACATGCCAGCGGATCATCGTGTACCGGATGCGTAGGAGACGGCCAGCTGAAAAACAC 174
QY 1877 TGGTGAAGTACGAGTGCACAGTTTTCAGTTTCTTAAATCCATTGTTAGAGGTACA 1936
Db 173 TGGTGAAGTACGAGTGCACAGTTTTCAGTTTCTTAAATCCATTGTTAGAGGTACA 114
QY 1937 ACCCTAGACTAAACGGTAATTGTTGTTGAAGAAAGAGTGAACACAGATTTTTCCTCAGT 1996
Db 113 ACCCTAGACTAAACGGTAATTGTTGTTGAAGAAAGAGTGAACACAGATTTTTCCTCAGT 54
QY 1997 CTGAGGAAGACTTCAGATCCACTTCCTGGAACAGTTATTGATGTAGAGTT 2049
Db 53 CTGAGGAAGACTTCAGATCCACTTCCTGGAACAGTTATTGATGTAGAGTT 1

```

RESULT 13

```

LOCUS BI560710
DEFINITION 603254744F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:5296954 5',
mRNA sequence.
ACCESSION BI560710
VERSION BI560710.1 GI:15448024
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 779)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11751 row: h column: 11

```


High quality sequence stop: 732.

FEATURES

source

Location/Qualifiers
 1. 779
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5296954"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_97"
 /note="Organ: testis; Vector: pBluescriptR (modified pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (Gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.2 kb and normalized to R0T 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation); Library constructed by M. Brownstein (NIMH/NIHRI, National Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 234 a 159 c 205 g 181 t
 ORIGIN

Query Match 17.4%; Score 405.6; DB 12; Length 779;
 Best Local Similarity 98.6%; Pred. No. 4.9e-97;
 Matches 430; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
 QY 8 TTGGTGTGAACACAGGAGCACTTAGACCATGTTTAAAGATCAAAAACAGGTTCTTCAG 67
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 344 TTGGTGTGAATACAGGAGCACTTAGACCATGTTTAAAGATCAAAAACAGGTTCTTCAG 403
 QY 68 GCATTATAGTAGGTAGGCACTAACCACTTCCGGCTGACATCCCGTCCCGAGGGGCT 127
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 404 GCATTATAGTAGGTAGGCACTAACCACTTCCGGCTGACATCCCGTCCCGAGGGGCT 463
 QY 128 TATATCAGTATCACTATCACTATACCACTGATGGAGCCAGAGCTCCGTTTCAGCTC 187
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 464 TATATCAGTATCACTATCACTATACCACTGATGGAGCCAGAGCTCCGTTTCAGCTC 523
 QY 188 TTCTTTTCAACAGAGATCAATTTGAAAGTGTGATGCTTTTGTGGAACGATATTAT 247
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 524 TTCTTTTCAACAGAGATCAATTTGAAAGTGTGATGCTTTTGTGGAACGATATTAT 583
 QY 248 TTTTACCTTAAAGACTACAGCAAGGTACTAGATTTTGTAGTAAGCCGGATGAG 307
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 584 TTTTACCTTAAAGACTACAGCAAGGTACTAGATTTTGTAGTAAGCCGGATGAG 643
 QY 308 AGGATGTGAGGATAACGATCACTTTTAAAGATGATGCTTTTGTGGAACGATATTAT 367
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 644 AGGATGTGAGGATAACGATCACTTTTAAAGATGATGCTTTTGTGGAACGATATTAT 702
 QY 368 TGCAGTTCTATATATTATTTTTCAGGAGGCTTTTGAATAATCA-TGAAATTTGCAACAATT 426
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 703 TGCAGTTCTATATATTATTTTTCAGGAGGCTTTTGAATAATCA-TGAAATTTGCAACAATT 762
 QY 427 GGAGCAATATTATTA 442
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 763 GGAGCAATATTATTA 778

RESULT 14

BU234973
 LOCUS
 DEFINITION 603411470F1 CSBQCHN24 Gallus gallus linear EST 26-NOV-2002 sequence.
 ACCESSION BU234973
 VERSION BU234973.1 GI:25479946
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 664)
 AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,

Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 12445392
 Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

Location/Qualifiers

source

1. 664
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST329K18"
 /dev_stage="22"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN24"
 /note="Organ: heads; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."
 BASE COUNT 182 a 132 c 162 g 188 t
 ORIGIN

Query Match 17.2%; Score 400.4; DB 13; Length 664;
 Best Local Similarity 76.5%; Pred. No. 1.1e-95;
 Matches 491; Conservative 0; Mismatches 151; Indels 0; Gaps 0;
 QY 1676 CAGATTGTTGCGAGCATCAATGAGGAGTATGACCCGCTGGTCTTCACGCTGCATATTC 1735
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 1 CTGGATTGTTGGCTAGCTGAATGAAATGACACGGTGGTTTTCACGCTGGCTTCTC 60
 QY 1736 AGGATAGAGCAGGAGCTGGTAGATGGGCTCAAGTCTGCTGCGAGCGGCTTGAGGG 1795
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 61 AAAGCCGTGGGAGGAAATTTGGATGGGCTCAAGCTGCTTGCAAACTGCCTTAAGGG 120
 QY 1796 CTTGGAATAGTCAATGAGTACATGCCAGCGCGATCATCGTGTATCCCGGATGGCGTAG 1855
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 121 AATGTTCAAGTGAATAAGTATTTGCCCTCTCGTATATTGTGTATCGTGATGGTGTAG 180
 QY 1856 GACACGCCAGCTGAAACACTGGTGAACTAAGAGTGCCACAGTTTTCGATGTGTATA 1915
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 181 GAGATGACAGCTCAATCTTTAGTGAACCTATGAAGTGGCTCAGTTTCTGATTTGTTGA 240
 QY 1916 AATCCATTGGTAGAGTTACACCCCTAGACTAACCGTAATTTGGTGAAGAAAGATGA 1975
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 241 AGACTTTGTTAAAGACTCAATCCAAAGACTGACTGTGATCGTTTGTGAAGAAACGAGTGA 300
 QY 1976 ACACCAAGATTTTTCCTCAGTCTGGAGGAGACTTCAGAAATCCACTTCTCTGGAACAGTTA 2035
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 301 GTACCAAGATTTCTTGGCAGGCTGGTGGAGGACTTAAACCCACCCCTTGTTGTTGTCG 360
 QY 2036 TTGATGTAGAGTTTACCAACAGAGATGTTATGACTTTTTTATCGTAGCCAGGCTGTGA 2095
 Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 361 TTGATATAGAGTGCACACAGCAGAGATGATGATTTCTTTTATTTGAGTCCAGGAGTGA 420
 QY 2096 GAAGTGGTAGTGTCTTCCACACATTCATGTCATCTATGACAAACAGCGGCTGAAGC 2155

